Age, sex and the life course: population variability in human ageing and implications for bioarchaeology

SHARMAN, JENNIFER,ALEXIA

Age, sex and the life course: population variability in human ageing and implications for bioarchaeology

Jennifer Sharman

Keywords: bioarchaeology, human skeletal ageing, sexual dimorphism, human variation

Abstract

Sex and age identification of human skeletal remains is essential in forensic anthropology, bioarchaeology and palaeodemography, and estimations rely on the use of proven methods. Many methods exist and are generally applied to skeletons from all time periods and geographic locations, despite studies suggesting that there are differences in the expression of traits characteristic of males and females and that ageing rates vary within and between populations.

The aim of this project was to study variation in ageing and sexual dimorphism in six documented collections from different geographic locations and/or time periods. Age and sex methods were tested on adult skeletal remains dating from the 17th to 20th century from Canada, England, South Africa, and Portugal. Ageing methods used were focused on the fourth rib's sternal end, cranial sutures, pubic symphysis and auricular surface. A more subjective age estimate for each individual was also produced, using informal skeletal age indicators alongside formal methods. Sex determinations were based on pelvic and skull morphology, and metrical analysis.

Differences were found between some collections in terms of the distribution of age phases and mean ages per phase. Similarly, distributions of sexually dimorphic traits were found to differ between some of the collections. In terms of overall age estimates, the subjective age estimates were significantly better than estimates based only on formal ageing methods, and intraobserver error tests suggest that user experience was important. The magnitude of such differences and their implications for bioarchaeology, forensic anthropology and palaeodemography are discussed.

Age, Sex and the Life Course: Population Variability in Human Ageing and Implications for Bioarchaeology

Jennifer A. Sharman

Submitted to fulfill the requirements of Doctor of Philosophy

Department of Archaeology

University of Durham

2013

Table of Contents

Acknowledgements

I must first thank my supervisors, Professor Charlotte Roberts and Dr. Becky Gowland, for their dedicated academic support and advice. Their comments on many abstracts, chapter drafts and presentations based on this research have improved this work enormously. Both Charlotte and Becky have supported me with patience and kindness, not only in academic crises, but personal crises as well, and for this I am very, very grateful.

I would also like to thank all of the people in charge of the collections that I visited, and many others who I met in travels to these collections. I would like to thank Mr. Roger Bulgin and Dr. Max Friesen from the Department of Anthropology, at the University of Toronto, in Toronto, Canada, for their permission to use the Grant Collection. From the Natural History Museum, in London, UK, many thanks to Mr. Rob Kruszynski and Dr. Margaret Clegg, for their permission to use the Spitalfields Named Sample, and their help while I was there. From Portugal, I would like to thank Dr. Hugo Cardoso for his permission to use the Lisbon Collection, and Dr. Ana Luisa Santos for her permission to use the Coimbra Collection, and to both for their hospitality and kindness while I was in Portugal. Many thanks to Mr. Jason Hemingway and Mr. Brendon Billings from the University of the Witswatersrand, Johannesburg, South Africa, for their permission to use the Dart Collection, for their advice on where I could (and could not) go around the university, for ferrying me around and for their hospitality. Thanks also to Dr. Victoria Gibbon-Woodward for her company and advice while I was working on the Dart Collection, and all the others who were so kind and hospitable from the university. I would also like to thank Dr. Ericka L'Abbe and Mr. Marius Loots from the University of Pretoria in South Africa, for their permission to use the Pretoria Bone Collection. Many thanks as well to Ericka, Marius and Jolandie Myburgh for their hospitality while I was in Pretoria, for showing me the sights and for feeding me! Thanks also Dr. Steve Symes and Susanne Daly for the great company in Pretoria.

Many thanks also to Dr. John Albanese, of the University of Windsor, for his help, advice, encouragement and the spreadsheets with formulae for his metrical method – they made the calculations much faster! Thank you also to my friend Ms. Hope Kron, for her help in Toronto at the Grant Collection.

I would also like to thank my colleagues and friends at the University of Durham and the University of Western Ontario for their support, advice, and osteological conversations.

I must also thank my mom, all of my family and friends, in Canada and the UK, for their support – their encouragement from afar was especially important when I wasn't able to get back to Canada as much as I would have liked. My special thanks to my dad, Chris and Jacqui – this has been much easier thanks to them.

Certainly not least of all, I thank my husband Matt, for his encouragement, support, patience, kindness, for dog-and cat-sitting when I was collecting data abroad for months, and for telling me to get on with it when I needed telling. It has been a long and sometimes difficult four (plus) years, and I thank you for being there with me for all of it. I am so grateful that you have been, and am looking forward to our next challenge! I promise that I won't do any more university degrees after this.

List of Tables

Appendix 8: Mean Age per Phase/Score

Appendix 18: Raw Data for Lateral-Anterior Cranial Suture Closure Age Estimation Method

Table A18.1. Raw data for lateral-anterior cranial suture closure method **(CD)**

Appendix 19: Raw Data for Vault Cranial Suture Closure Age Estimation Method

Table A19.1. Raw data for vault cranial suture closure method **(CD)**

Appendix 20: Raw Data for Fourth Rib Method

Table A20.1. Raw data for fourth rib method **(CD)**

Appendix 21: Raw Data for Summary Age Estimates

Table A21.1. Raw data for summary age estimation methods **(CD)**

List of Figures

‘While I have felt justified in speaking with greater confidence on the determination of the sex than I did sixteen years ago, in the question of age it is just the reverse.’

(Dwight, 1894)

‘Stay passenger stay and read this stone, Remember how soon we were gone
Death does not always warning give; Dear friends be careful how you live.’

(Inscription on the gravestone of George, Sarah, Ann and Mary Ayre,
All Saints Parish Church, Lanchester, County Durham, UK)

Chapter 1: Introduction

1.1 Research Context

Estimates of age-at-death and sex from adult human skeletal remains are fundamental to interpretations of health, demography and social identity in the past and for forensic anthropology. The ageing process and sexual dimorphism in the adult human skeleton vary over time and space, depending on environment, genes, activity, functional differences, socioeconomic status, and psychosocial factors, with interaction between these factors, affecting population and individual level variation (Borkan and Norris, 1980; İşcan and Loth, 1989: 36; Sherman, 1999: 11; Molleson, 1986: 106; Rowe and Kahn, 1987: 144, 148; Kobyliansky et al., 1995: 88; Sayer et al., 1998; Cox, 2000: 63-64; Buckberry and Chamberlain, 2002). These include:

- socio-political circumstances that may affect a population's rate of ageing resulting in secular differences (Albanese, 2003b: 174-175);
- differences in the intensity and level of physical activity (occupational and/or leisure) within an individual's lifetime that can lead to variation in ageing rate over the life course (Karasik et al., 2005: 578), on an individual and at the population level;
- pathological conditions or trauma that may affect the morphological expression of age-related joint degeneration (Klepinger et al., 1992: 768);
- differences in masticatory stress due to the level of dietary coarseness or cultural practices, such as the use of teeth as tools, that may lead to variation in craniofacial robusticity, thus affecting sexual dimorphism (Jantz and Meadows Jantz, 2000: 335; Mays and Cox, 2000: 125); and
- environmental factors and change in demographic parameters that may result in secular change in craniofacial morphology (Jantz and Meadows Jantz, 2000: 335; Buretić-Tomljanović et al., 2006: 674; Gonzalez et al., 2010: 377; Weisensee and Jantz, 2011: 556-557), and these may also lead to population differences in sexual dimorphism.

There may also be biological sex-related differences in morphological ageing for some joints; for example, in the sacroiliac joint due to mechanical/functional differences (Stewart, 1984: 195). Despite this variation, the methods used in bioarchaeology and forensic anthropology to determine age and sex of individuals from their skeletal remains are typically developed using one skeletal sample but the methods will only reflect variation within that sample. The methods are then typically applied to other skeletal samples, regardless of those individuals' origins.

Either implicitly or explicitly, applying these methods globally relies on uniformitarian theory (Howell, 1976). Uniformitarian theory states that biological processes, including how a person's skeleton ages, have not changed over the course of human history (Howell, 1976: 26;

Paine, 1997; Chamberlain, 2000), which is the assumption that researchers make when applying standard osteological age and sex determination methods to archaeological skeletal remains. However, skeletal indicators of age-at-death are notoriously unreliable and frequently over-estimate the age-at-death of young adults and under-estimate those of older adults (Meindl et al., 1983; Bedford et al., 1993; Cox, 2000). New Bayesian statistical approaches have attempted to deal with these inaccuracies (Hoppa and Vaupel, 2002; Chamberlain, 2006), but do not fully address the fundamental problem: lack of clarity of the scope of interpopulation variability in human ageing and sexual dimorphism.

1.2 Aims and Hypothesis Tested

The purpose of this dissertation was to investigate the hypothesis that ageing rates and sexual dimorphism vary across adult human populations, to the extent that current skeletal methods used to estimate age and sex may be inappropriate for some populations. To do this, I tested the null hypothesis that there is no variation in human skeletal ageing rates or expression of sexual dimorphism between populations from different geographic locations or time periods, and the current skeletal methods used can be applied indiscriminately to all populations.

The research had two aims:

- To quantify variability in skeletal ageing and sexual dimorphism by analysing geographically and temporally diverse skeletal populations of known age and sex and testing the efficacy of existing methods.
- To assess the relationship between skeletal age indicators and sexual dimorphism over the life course.

While evidence for variability in ageing rates both within and between individuals, between the sexes, and within and between populations has been found (Kemkes-Grottenthaler, 2002; Sherman, 1999; Ruff and Hayes, 1982; Suchey, 1979; Katz and Suchey, 1986; Molleson, 1995), the extent of such variation is not yet clear. This research has the objective to clarify the nature of the variation in ageing rates and sexual dimorphism between populations.

1.3 The Study

This research tested the efficacy of the skeletal indicators typically used to estimate adult age and sex; this was an exploration of variability in ageing rates and expression of sexual dimorphism, because these are amongst the most important reasons that skeletal indicators developed on one documented (known age-at-death and sex) population may not be applicable to undocumented populations, separated by time or space. As the data were collected from the Grant Collection (early 20th C, Toronto, Canada), Christ Church, Spitalfields (18th to 19th C, London, England), Lisbon and Coimbra Collections (early 20th C, Portugal), and Dart and Pretoria Collections (20th C, Johannesburg and Pretoria, South Africa) (see Bedford et al., 1993; Molleson and Cox, 1993; Cardoso, 2006; Albanese et al., 2005; Dayal et al., 2009; L'Abbé et al., 2005), within- and between-country data comparisons were possible.

The Christ Church, Spitalfields collection was the earliest sample analysed, but it allowed for comparison between samples from different time periods, as the other collections are more recent in date. While much work has been done using the Christ Church, Spitalfields collection and the Coimbra Collection, less research has used the other collections. The South African collections have been used mainly by South African researchers (e.g. Steyn and İşcan, 1997; Asala, 2001; Pretorius et al., 2006; Oettlé et al., 2009), and the Lisbon Collection has only recently become available for research. These under-utilized collections added an interesting dimension to this research, and it could be argued that they avoided bias in terms of prior knowledge or assumptions regarding the collections.

Biological anthropological and forensic anthropological studies have tested the efficacy of skeletal methods of age and sex determination, but typically use only one sample (Saunders et al., 1992; Bedford et al., 1993; Falys et al., 2006; Hens et al., 2008; Williams and Rogers, 2006), or archaeological samples, in which age or sex is first estimated with one method, against which the results of a second method are compared (Dreier, 1994; Gillett, 1991; Nagaoka and Hirata, 2008; Oliveira et al., 2006; Walrath et al., 2004). However, by their very nature, archaeological samples cannot verify or refute the accuracy of age and sex estimation methods, as these are unknown parameters in such populations. While tests of methods using only one sample are useful, they cannot truly add to discussions on interpopulation variability in age and sex determination accuracy, even if compared to results of other studies, because differences between the ages and sexes obtained could be artifacts of intra- and interobserver error and not variability in ageing or sexual dimorphism. In this research, because the author collected all data from all collections sampled, interobserver error was not an issue (although both inter- and intraobserver error were tested); thus, any differences found reflect variation in ageing rates and sexual dimorphism.

The word "ageing" is used throughout this thesis in several different ways. Skeletal ageing refers to the estimation of age-at-death of human skeletal remains. Of course, the skeleton (and the individual) is also subject to ageing in the sense of the degeneration that occurs with age, of senescence. There are also ageing populations in the demographic sense, with a rapidly expanding proportion of older adults, a common situation in much of the modern world.

1.4 The Challenges

It is well known that the current skeletal ageing methods do not identify ages-at-death in the oldest age categories; some argue that this was a biological reality for archaeological populations, i.e. that people simply did not live that long. Archaeological samples with ages-at-death estimated from skeletal indicators thus do not usually show very many old people – often open ended categories are given (such as “50 plus” or “60 plus”; e.g. Lovejoy et al., 1977; Mensforth, 1990; Alesan et al., 1999) after combining estimates from more than one ageing method that each provided varying age ranges. However, another possibility is that the absence of the “oldest old” people in archaeological samples is an artifact of the ageing methods (Gowland, 2007: 156). While it is likely there were fewer people living into very old ages in the past (due to the lack of modern medicine, and other relatively recent advances in sanitation, etc.), it would be odd if none did; indeed, some studies using ancient documentary evidence have in fact noted survival into old age (i.e. over 75 years of age) in the past (Montagu, 1994: 25-26; Zhao, 1995: 99). This research has the opportunity to also explore this problem – skeletons of the “oldest old” ages-at-death are present in these documented collections, and were analysed to examine the amount of error involved in estimating old ages.

While documented collections of human skeletal remains are invaluable for developing age and sex estimation methods that are used in bioarchaeology and forensic anthropology (Usher, 2002), they are not without problems. Such collections are not typically representative of the population from which the individuals originate and, when ageing methods are developed on unrepresentative populations, a potential pitfall, termed “age mimicry” by Bocquet-Appel and Masset (1982), can occur when the methods are applied to other skeletal samples. Age mimicry occurs when the unknown sample (the target sample) is aged using a particular skeletal indicator, with subsequent age-at-death distributions tending to mimic those of the collection used to develop the method (the reference sample). For example, if an archaeological sample of skeletons was aged solely using McKern and Stewart’s (1957) pubic symphysis method, developed on sample of young American war dead, the estimated age-at-death distribution might show more young males than were actually in that (sample) population.

Many solutions have been presented to remove the effect of age mimicry, including using multiple methods of age estimation and Bayesian statistical methods (Konigsberg and Frankenberg, 1992; Kemkes-Grottenthaler, 2002; Hoppa and Vaupel, 2002; Chamberlain, 2006). While these are viable and useful contributions to the problem, it is important to gain an increased understanding of the variability of expression of observed features that are used to assess age-at-death and sex between and within populations, and this is the overall purpose of this dissertation. It is the contention of the author that, while statistical methods used for analysing the data are important, it is equally or more important to examine the reasons behind variability in ageing and sexual dimorphism, and to understand the breadth of this variability. Only when we have a better and more thorough understanding of the issues we are dealing with will we be able to offer a possible solution for bioarchaeology and forensic anthropology.

Furthering the problems of age estimation from human skeletal remains is the fact that it is unclear why and how ageing actually occurs, although some good candidate hypotheses include antagonistic pleiotropy and mutation accumulation (Crews, 2003). Perhaps the turnover of bone that slows with age provides an example of mutation accumulation, while failure of organs in advanced age could be an example of antagonistic pleiotropy. This is a simplified example, of course, but these theories deserve elucidation.

These problems and challenges will be explored in relation to human skeletal ageing, and the potential effects that are undermining our efforts to estimate age using skeletal indicators.

1.5 Implications

This research will identify and qualify the variability in rates of ageing and expression of sexual dimorphism within and between skeletal populations from differing geographic and temporal locations. Quantification of levels of error and bias associated with current age and sex determination techniques has not been undertaken previously on such a scale. This work will make significant contributions to clarifying the nature of variation in the ageing process and sexual dimorphism in living and past populations, and increasing the accuracy of age and sex estimation methods using skeletal remains. It will also have broader implications for forensic anthropology (biometrics and human identification), paleodemography, and clinical studies of ageing.

As many developed countries in the world are experiencing increasingly ageing populations, with more people surviving into their eighties and nineties (Audit Commission, 2008; Beard et al., 2012: 5), this work also has the potential to provide important information on variability in ageing for health care professionals and policy makers. Indeed, research on the variability in ageing in living humans and “successful” ageing (as opposed to “unsuccessful” ageing, with a high cost to health care and other public services) is abundant in the literature (e.g. Litwin and Shiovitz-Ezra, 2006; Kaplan et al., 1987; Bassuk et al., 2002; Landi et al., 2007; Rowe and Kahn, 1987; Beard et al., 2012). By highlighting differences in rates of ageing in populations from different places and time periods, and by analysing the parameters involved in such variability, this work will contribute to advancement of knowledge in modern and ancient contexts.

1.6 Structure

Chapter 2 outlines the development of the currently used age and sex determination methods, as well as their limitations. The use of documented collections is also discussed, as are the theories of human senescence, and there is a brief discussion of human variation. Chapter 3, Materials and Methods, provides a discussion of the documented collections sampled in this research and their historical contexts, as well as the specific age and sex determination methods being tested. The statistical methods used in data analysis are also outlined. Chapter 4 provides the results of the data analysis, including inter- and intraobserver error tests, while Chapter 5 discusses the results with respect to their implications for bioarchaeology and forensic anthropology. Finally, Chapter 6 summarises the research, including limitations of the study and some suggestions for future research.

Chapter 2: Ageing and Sexual Dimorphism in Bioarchaeology and Forensic Anthropology

2.1 Introduction – Importance of Age and Sex Estimation in Bioarchaeology and Forensic Anthropology

Meindl and Russell (1998:378) call sex 'the most important demographic variable'; age-at-death is certainly the second most important. While Meindl and Russell were commenting on the importance of sex determination for paleodemography, the same can be said for bioarchaeological studies generally. Most bioarchaeological studies require basic paleodemographic reconstruction, and accurate sex and age estimates are essential for this reconstruction and its analysis (Mensforth, 1990: 89; Konigsberg and Frankenberg, 1994: 101; Nagaoka et al., 2006: 8). In forensic anthropology, individual identification is very much reliant on the accurate determination of sex and age because these variables are included in missing person data. Many methods of age and sex estimation exist and are variously applied to skeletal remains from populations of all time periods, and from both archaeological sites and crime scenes all over the world.

However, studies show differences in the expression of sex characteristics, and populations from different locations and time periods may age at different rates due to cultural, environmental and biological differences (Buckberry and Chamberlain, 2002: 231). As outlined earlier, the purpose of this thesis is to investigate this variability in ageing rates and expression of sexual dimorphism, and their impact on accuracy of estimates. As such, further elucidation on the methods used to estimate age and sex is necessary here.

In this chapter, variability in human ageing rates and sexual dimorphism expression and the factors involved in such variation are discussed. The historical developments of the methods used to age and sex skeletal remains are described in this chapter, with a focus on the methods tested in this research. This follows a discussion of the documented collections that have been used to develop age and sex determination methods and the limitations of such collections. Senescence, ageing and human longevity will also be considered, as the reasons for ageing and senescence are important to evaluate when thinking about the causes of population variation. First, human variation and the reasons underlying this variation are discussed.

2.2 Human Variation

This section describes human variation, in sexual dimorphism and ageing rates, alongside the reasons underlying this variation. "Race" is also discussed in terms of variation, although studies using "race" as a variable may be overlooking the true causes of variation between populations.

As a species, *Homo sapiens* is adaptable, plastic and variable; one only has to think of superficial differences in skin colour, eye colour, body shape and height for examples. This plasticity, the ability to react to change in the environment and activity levels, coupled with the wide range of human habitats and activities (work and leisure) that humans experience, contributes to divergent ageing rates and sexual dimorphism expression among and even within populations (e.g. Carlson et al., 2007: 18; Walker, 2008: 48; Kemkes-Grottenthaler, 2002; Chamberlain, 2006: 105; Falys and Lewis, 2011: 705). Humans vary not only by phenotype, but also genotype; the interaction between genes, environment, culture and behaviour is complex, and works to shape an individual's phenotype over the lifespan. While there is, as mentioned above, genetic variation between humans, there is more genetic variation within any specific population than between populations, or groups referred to as a "race" (Lewontin, 1972; Brown and Armelagos, 2001; Relethford, 2002).

Variation, in phenotype and genes, occurs in clines, or continua, that do not necessarily (or perhaps even often) coincide with "racial" groupings (e.g. Chikhi et al., 1998; Handley et al., 2007). The phenotypes that may be considered indicative of membership to particular "races" are expressed clinally across geographic locations – there are no strict borders between "races" or ethnic groups (Keita and Kittles, 1997: 537; Relethford, 2009: 21-22). Race is a social-cultural construct, not a biological one (Keita and Kittles, 1997: 539); biological human variation does exist, but not very much of it is centred on the socially identified "racial" groups. Where there are biological correlates, there is evidence that the reason for the correlation may be socio-political in origin (Gravlee, 2009). It is true, however, that there are certain genes (or gene mutations) that are more prevalent in certain groups of people; these may affect frequencies of certain diseases in people with shared ancestry. For example, sickle cell anemia tends to be more common in African (or those of African descent) and southern Mediterranean inhabitants (Polednak, 1989). Another example is Tay-Sachs disease, a genetic disorder found in Ashkenazi Jewish groups. There are other non-Jewish groups with a similar prevalence of this disease, but as a result of different mutational variants (Myerowitz and Hogikyan, 1986: 1648). That this is a "Jewish disease" is a result of genes and culture – Ashkenazi Jews historically tended to marry and reproduce with other Ashkenazi Jews.

2.2.1 "Race" and Variation

"Race" was much discussed by early biological anthropologists (e.g. Hrdlička, 1918; Hooton, 1918), and, unfortunately, unflattering descriptions were often used for non-white, non-Western groups of people, to insinuate a ranking of "races" (e.g. Coon, 1963). While many researchers do not agree with the concept of "race" as a strict biological entity now (e.g. Relethford, 2002, 2009; Gravlee, 2009), and certainly the unflattering descriptions and rankings are generally not present in current bioarchaeological and forensic anthropological research, groups of people are often still categorised in the same way as in the earlier research (e.g. Gill and Rhine, 1990; Patriquin et al., 2003). Such categorisation may disguise true (normal) variation (e.g. Stevenson et al., 2009: 439). This can be problematic in the interpretation of data – for example, in the development of sex determination methods. Albanese (2003a: 3) realised this and, in constructing a sample for a metrical method of sex determination, controlled for age-at-death and year of birth instead of "race", knowing that there is variation in the expression of sexual dimorphism with age (Walker, 1995, 2008). The Terry Collection used for this study had a higher number of young white females collected late in the collection's history to make up for underrepresentation of this group; older black females were collected earlier in time. Indeed, Albanese (2003a: 3) could predict "race" with an accuracy of 69.5% with logistic regression using only the variables of year of birth and age-at-death. If "race" had been controlled for, the sample would likely have included a disproportionate number of old black females and young white females, leading to distortions in the models formulated due to age differences in sex trait expression. The same was found to hold true with regard to secular change in femur length, used to illustrate this problem – while significant differences were found by birth cohort, no significant "race" differences were found when birth cohort was the controlled variable (Albanese, 2003b: 174-175). Thus it is important to examine sources of human variation other than "race".

2.2.2 Variation in Sexual Dimorphism

This thesis is concerned specifically with variability in skeletal ageing rates and sexual dimorphism. All human populations are skeletally sexually dimorphic, but the degree or scale of sexual dimorphism can differ between populations (Kajanoja, 1966: 32; MacLaughlin and Bruce, 1990: 1391; MacLaughlin and Bruce, 1986b: 230; Steyn and İşcan, 1998). Variation in the degree of sexual dimorphism is when the difference between typical male and female morphology may be wider or narrower (e.g. MacLaughlin and Bruce, 1986; Carlson et al., 2007: 18; Walker, 2008: 48). Similarly, some populations may be more masculine or feminine overall – for example, if scoring a particular skeletal trait from 1 to 5, where 1 is very feminine and 5 is very masculine, the average score of males for one population might be a 4, and females a 2, while the average score of males from another population might be 2 and females a 1 (e.g. Walker, 2005: 387-389; Washburn, 1949: 428-429). If using metrical methods, sectioning points (the differentiating point between male and female) can vary by population and overlap, illustrating variation in scale or range of sexual dimorphism (Rosenberg, 2002: 13).

2.2.3 Variation in Ageing Rates

Variation in ageing rates was recognised as early as the 19th century. Pommerol (1869: 26) stated that any estimation of age from a skull was just that – an estimation, conjecture, as a precise estimate was impossible. It was recognised that variation was wide, with many different causes, differences between populations and "races", and between people who lived in the same geographic location (Pommerol, 1869: 26-27). Many studies have found evidence of variation in ageing rates or expression of age-related morphological change between populations that vary in time and geographic location (Genoves, 1960: 206; Semine and Damon, 1975; Katz and Suchey, 1989; Hoppa, 2000; Oettlé and Steyn, 2000; Schmitt et al., 2002; Schmitt, 2004), but these studies tend to compare only two samples. Others have found satisfactory results in applying an age indicator on a population not related and quite far geographically from the population used to develop the method (Yavuz et al., 1998), but this could simply be coincidentally similar ageing rates for that particular age indicator. Interestingly, Schmitt et al. (2002: 5) found patterns of ageing in samples geographically closer to be more similar to each other than between samples geographically far from each other; perhaps clinal variation of sorts may be seen in ageing rates. Further, interobserver error can be significant in some methods of age and sex estimation, and observer experience can also affect the accuracy of a method (e.g. Walrath et al., 2004), although others have found only non-significant levels of interobserver error (Galera et al., 1995). As such, results of studies with different observers are not strictly comparable.

2.2.4 Factors Underlying Variation

The reasons for variation in ageing rates are many, including diet, environment, occupation, activity level, functional differences, socioeconomic status, psychosocial factors, and genes, with interaction between factors, with not only population variation but also individual variation (Borkan and Norris, 1980; İşcan and Loth, 1989: 36; Sherman, 1999: 11; Molleson, 1986: 106; Rowe and Kahn, 1987: 144, 148; Kobyliansky et al., 1995: 88; Livshits et al., 1996: 551; Sayer et al., 1998; Cox, 2000: 63-64; Weiss and Jurmain, 2007: 443, 445; Sahni et al., 2010; Campanacho et al., 2012: 375). Interaction between factors may include intrapopulation differences in socioeconomic status resulting in variation in diet quality or quantity; how differential dietary quality affects individuals is also dependent on individual differences in environment, genes and psychosocial factors. Variables may also change over the life course of an individual: for example, activity level may diminish with age (Karasik et al., 2005: 578), and dietary requirements change with age. Environment, socioeconomic status, psychosocial factors and functional differences may also change over the life course. For skeletal ageing at the individual level, variation in circulating concentrations of sex steroids and parathyroid hormones with age may also have an influence on bone density and bone turnover, respectively (LeBoff and Glowacki, 1999: 159;

Silverberg and Bilezikian, 1999: 175). Estimates of the contribution of inherited factors responsible for determining bone density have varied somewhat, between about 50 and 70%, though Krall and Dawson-Hughes (1993: 2, 8) estimated heredity to be responsible for 46 to 62% of the interfamilial variation in bone density for their sample; environmental factors could thus be responsible for half of variation. Ancestral population differences have also been reported for bone density; black individuals (living in Africa or North America) tend to have higher bone density for age than white (North American or European) or Asian individuals (Broman et al., 1958: 210; Trotter et al., 1959: 25; Stini, 1994: 155).

The factors behind variation in sexual dimorphism have not been considered as widely, but may include environment (Borgognini Tarli and Repetto, 1986), gendered allocation of labour (Ruff, 1987; Bridges, 1989; Pomeroy and Zakrzewski, 2009), stress (environmental, nutritional or disease, with a decrease in sexual dimorphism suggested when due to nutritional deficiency; Stini, 1969; Gray and Wolfe, 1980), socioeconomic status (Steyn and İşcan, 1999: 83), and genes (Rice, 1984; Borgognini Tarli and Repetto, 1986). Secular change, in some cases over a remarkably short time period, have been reported – in sexual dimorphism (e.g. a decrease over time, İşcan et al., 1995), in craniofacial traits (Smith et al., 1986; Jantz and Meadows Jantz, 2000), which may have implications for sex determination, as some traits changed in females but not males (Buretić-Tomijanović et al., 2006), and in height and weight (Eveleth et al., 1979; Garn, 1987; Tobias, 1988). Some evidence has also suggested that females are less sensitive to environmental stress than males, although other studies provide contradictory evidence (e.g. Stini, 1969; Moore et al., 1998: 54; Stinson, 1985; Sheridan and Van Gerven, 1997: 251), a phenomenon called female buffering. It is thought to be conferred on females due to higher investment in reproduction – the buffering is necessary to protect against the stress of pregnancy and lactation (Stinson, 1985: 123).

Sexual dimorphism can vary even across a relatively small geographic area (e.g. Rosenberg, 2002; Cunha and van Vark, 1991). Cunha and van Vark (1991: 63) found variation within the southern, central, and northern regions of Portugal. Underlying these factors is historical and politico-economic context, which should also be considered when assessing interpopulation variation. Because the reasons underlying differences in ageing rates and sexual dimorphism are complex and interactive (Frayer and Wolpoff, 1985: 432), the relative contribution of each of these is not known and is likely to differ by population. As there is variation through space, there is no reason to think that there would not also be variation through time – however, this is more difficult to test, as it would require known age populations from different (but distinct) time periods without overlap, but from the same geographic location, preferably with known population continuity.

2.2.5 Levels of Variation

There is also variation at two other levels: between individuals and within individuals (Cox, 2000: 64). Interindividual variation can occur for reasons similar to population-level variation: genes, environment, and culture (including behaviour) (Baccino and Schmitt, 2006: 261). There is evidence that pathological conditions can affect ageing rates, and some have suggested that alcoholism and drug use could equally be implicated (Katz and Suchey, 1989: 172), as alcoholism can lead to a decrease in bone mass and formation (Saville, 1965; Turner, 2000; Santori et al., 2008). Klepinger et al. (1992: 768), using the Suchey-Brooks's pubic symphysis age indicator method, estimated the ages of a few individuals who had suffered from long-term illness, presumably with low levels of physical activity, and an amputee (with resulting change in weight-bearing and biomechanical factors) – the ages of these individuals were all significantly under-estimated, suggesting that low activity levels affected the rate of degeneration of the pubic symphysis. Other morbidities have been associated with advanced skeletal ageing (here defined by the radiographic presence of osteophytes and tufting in the hand phalanges), including rheumatic diseases, ischaemic heart and pulmonary diseases (Kalichman et al., 2006: 81-82). The variation in ageing within individuals, between physiological systems or even between different skeletal elements, has been noted by others (e.g. Borkan, 1986: 85-87; Kemkes-Grottenthaler, 2002: 48; Spirduso, 1995: 37; Loth and İşcan, 1989: 118) and Wittwer-Backofen et al. (2008: 390) showed that age estimates for particular individuals ranged from 20 to 60 years at death, depending on the age indicator and observer. Intra-individual variation can be affected by functional differences, behaviour, occupation/activity, environment and genes (Cox, 2000: 64). As İşcan and Loth (1989: 36) note, '…each bone is only a single aspect of the skeleton, and, by its nature, has a different function from all others… [T]hese functional differences no doubt affect the manifestation of age.'

2.2.6 Variation and the Environment

The plasticity of the human skeleton has also been evidenced, showing its ability to adapt to extreme environments or in response to environmental change (where "environment" encompasses social, economic and political environment). For example, a study of Mayan immigrant children in the US showed greater stature and weight for age than their Guatemalan counterparts (although still slightly lower than American children), suggested to be the result of better nutrition and environmental conditions in the United States (Bogin and Loucky, 1997: 21, 27). This is in concordance with the above-mentioned studies of secular change: better nutrition leads to larger bodies, but a reversal can also occur, where deprivation leads to smaller bodies and/or more specific skeletal changes such as a decrease in certain pelvic diameters (Nicholson, 1945; Garn, 1987; Tobias, 1988). Evidence has shown that successive generations of children born to migrants continue this increase in height, weight, fat and musculature to eventually match that of the host population (Bogin and Loucky, 1997: 29). Skeletal development and growth are more sensitive to environmental insults than dental development (Lasker, 1969: 1485; Roberts, 1981: 244; Conceição and Cardoso, 2011: 469). As there is evidence of degree of sexual dimorphism reducing under nutritional stress (Stini, 1969), it might be expected that the degree of sexual dimorphism might also change from that of immigrant populations to their successive generations living in the "new" country with better nutrition; if height and weight are considered (Bogin and Loucky, 1997), this type of secular effect on sexual dimorphism might not take more than a generation or two to be visible.

Adaptation to extreme environments also occurs – for example, in people who live at high altitudes. Migrants (or visitors) to high altitudes will slowly acclimatise (or not), but those with long-term residence at high altitudes have adapted better, and perform better than those newly-acclimatised in measures relating to exercise capacity, have less hypoxic vasoconstriction, lower pulmonary arterial pressure and lower haemoglobin concentration (Moore et al., 1998: 56). Interestingly, high-altitude residents with a longer generational history of residing in that habitat seem to have adapted even better than lifelong residents who have historically fewer generations who have lived at high altitudes (Moore et al., 1998: 56).

2.3 Human Longevity

Modern estimates indicate a hypothetical maximum potential lifespan of around 120 to 125 years for *Homo sapiens* (Schulz-Aellen, 1997: 44; Weon and Je, 2009: 65; Crosse, 2011: 193). This proposed estimate is based on observations of the oldest humans (for example, Jeanne Calment, who died at the age of 122.5 years; Weon and Je, 2009: 65) and the fact that maximum lifespan and the ageing process are under genetic regulation, although rate of ageing is a product of interactions with other factors, such as environment and lifestyle (Ostojic et al, 2009: 687; Ricklefs, 2008: 380; de Magalhães et al., 2012). Perhaps the first confirmed centenarian was Eilif Philipsen, of Norway, who died in 1785 at the age of 103; validation and confirmation of centenarians before this date are lacking, but some other likely centenarians emerged shortly after (Kjæraard, 1995: 53-54; Hynes, 1995: 88).

The attraction of the middle, discussed earlier, has contributed to the idea that past populations were not particularly long-lived. Indeed, some paleodemographic reports use truncated age-at-death distributions ending with open-ended categories of 45+, 50 +, or 60+ as a result of this idea (e.g. Lovejoy et al., 1977; Mensforth, 1990; Alesan et al., 1999) and a recent literature review of bioarchaeological studies published between 2004 and 2009 has indicated that the oldest age category presented ended at age 70 (Falys and Lewis, 2011: 709). However, it is likely that lifespans were just as variable in the past as they are today (Jackes, 2000: 418). It may be true that the proportions of the oldest old were not as high as they are today, as there is good documentation for increased survivorship of those over 80 years old over the last few decades (e.g. Audit Commission, 2008: 16), but it seems highly unlikely that no one in the past lived beyond the age of 75 or 80 years, particularly in light of the known problems with estimating age at the oldest ages. Indeed, Cox (2000: 62) notes that the supposedly short lives of past populations (as reported in archaeological studies) are a product of the limitations of the skeletal ageing techniques.

During the medieval period (using records from 1276 to 1450), the average age-at-death for males over 15 in England ranged from 47 to 54 years at death (Jackes, 2000: 418-419), suggesting that survival above these ages did occur, as these are averages. In a study of recorded Greek and Roman birth and death dates of males living prior to 100 BC, Montagu (1994: 25-26) suggested that the average length of life for these males was the same as modern populations prior to AD 1950; a median of 72 years at death was calculated, comparable to 71.5 and 71 years for those who died between AD 1900 to 1949 and AD 1849 to 1899, respectively. While the anecdotal evidence for individuals living to 150, 172 and even 185 years at death (Easton, 1799: 14) seem unlikely given the postulated maximum length of human lifespan of 120 to 125 years (e.g. Schulz-Aellen, 1997: 44; Weon and Je, 2009: 65; Crosse, 2011: 193), the evidence available does seem to indicate that it is equally unlikely that the majority of past people did not live past 50 years of age, as Crews (2003: 1) suggests. Interestingly, Easton (1799: 15) quotes a contemporaneous estimation that only 6% of people in Europe live beyond 60 years; however, this small proportion could also partly be an artifact of smaller population size, as mortality and population size “interact” to dictate the actual number of people surviving past a certain age (Zhao, 1995: 93).

Using Chinese genealogy records for the Wang surname and beginning from around 500 BC, Zhao (1995: 97-98) analysed lifespan information; however, while data for some 30 000 people were recorded, females, those who died very young, and people who could not be traced were omitted and/or underrepresented, so the actual population size was not known. Data from 2 500 males with more reliable and complete information suggested that even before AD 1500, approximately 105 males (or 4%) survived to age 80 or more (Zhao, 1995: 99). For example, a centenarian was reported to have died in 1513; as these records were kept for the higher social classes, that the records of birth and death dates compared to age-at-death seem consistent and reliable, and given China’s large population compared to European countries, Zhao (1995: 100-101) suggests that a centenarian was not unreasonable. While not all individuals have recorded birth and death dates, because the time of birth was regarded as an important factor in fortune and destiny, they were likely recorded precisely and reliably; furthermore, no age heaping was found in the records of dates and ages-at-death, supporting the reliability of the available data (Zhao, 1995: 102).

2.4 Senescence

Scientists are still working to understand how and why we age. The human ageing process (or "senescence") is not caused by just one biological mechanism; alongside biological factors, environment and culture complicate descriptions of the process (Crews, 2003: 2; Ostojic et al, 2009: 687; Ricklefs, 2008: 380; de Magalhães et al., 2012). Senescence is the process of decline of physical and physiological functioning that progresses with chronological age in mammals, resulting in an increasing probability of death (Borkan et al., 1982: 182; Kirkwood, 1997: 684). There are currently some good candidate evolutionary theories to explain why we age, including antagonistic pleiotropy, age-specific gene action or mutation accumulation, the disposable soma and the thrifty or pleiotropic gene model (Crews, 2003). There are also a number of hypotheses to explain the mechanisms of ageing – including telomere shortening (Cameron and Demerath, 2002: 178), oxidative damage or stress (Finkel and Holbrook, 2000), and a number of mechanisms involving the gradual accumulation of damage to cells, proteins or DNA (Crews, 2003: 18; Ricklefs, 2008). It seems likely that a number of these theories are correct, and may account for different aspects of ageing and perhaps eventually for differences in rates of ageing, depending on the system and function under study (Crews, 2003: 247; Ricklefs, 2008). Indeed, as the force of natural selection declines with age and is more ineffectual at the oldest ages, no one theory is solely responsible for the decline in physiological functioning that results in death (Finch and Kirkwood, 2000: 66). Various types of cells, organs, and systems provide evidence for different theories.

This evolutionary perspective is particularly important, as ageing seems disadvantageous from the perspective of individual fitness. However, the influence of natural selection declines progressively with increasing age, as there are fewer individuals on which it may act – that is, discriminate between genotypes conferring more or less fitness (Kirkwood, 1997: 685; Crews, 2003: 14-15). This means that genes have a decreasing influence on rates of ageing with increasing age, and a contrasting increasing influence comes from the environment (lifestyle, cultural, behavioural). Supporting this is increased heterogeneity in physiological function, but a general decline with age (Kirkwood, 1997: 689). Some wonder whether humans must age, or whether, with advances in medicine and science, humans have the potential to live forever; others believe the evidence points to a genetically-regulated predetermined maximum human lifespan, with only the rate of ageing affected by environmental and other variables (Ostojić et al., 2009: 687). More recently, scientists have begun treating senescence as another phenotypic variant (Crews, 2003: 5-6). That is, ageing is a natural phenomenon, with much individual variation, but not so much variation that it cannot be understood. That senescence is affected by genes and environment, and is affected by more than one genetic factor, with no evidence suggesting a particular genetic program, indicates that studying senescence as a phenotypic variant is valid (Crews, 2003: 6, 240; Bell et al., 2012).

Antagonistic pleiotropy suggests that some genes can confer an advantage in early life or fitness (in reproduction, for instance), but can be disadvantageous later in life (Kirkwood, 1997: 685; Crews, 2003: 15-16). As natural selection has a stronger influence over the earlier portion of the lifespan, genes with an early benefit are selected for, with the benefits outweighing the later disadvantages (Crews, 2003: 38; Hughes, 2010: 1273). The levels or activity of testosterone in men is suggested to be a possible example of antagonistic pleiotropy. During puberty, testosterone promotes masculinisation of secondary sexual traits, and later promotes mitosis and growth of cells in prostate and muscle tissues; however, at the highest levels and at older ages, testosterone may increase the risk of cardiovascular disease, prostate cancer and hypertrophy (Crews, 2003: 39; Suekoa et al., 2010).

The **disposable soma theory** is related to maximisation of fitness (reproducing the germ line), while minimising the effort required to keep an organism fit (maintaining the soma), with all the necessary repairs (Kirkwood and Holliday, 1986: 6-7; Kirkwood, 1997: 685; Crews, 2003: 17; Lorenzini et al., 2011: 3853). Even in a non-senescent organism, accidents occur that would eventually result in death of all individuals at some point, resulting in the loss of investment in maintaining the individual. This means that the investment is maximised by maintaining the individual long enough for reproduction, but not longer than the expected lifespan, as this would be too high an investment and a waste of resources (Kirkwood and Holliday, 1986: 6-7; Kirkwood, 1997: 685). Under this theory, the random accumulation of damage is supported as a mechanism of ageing, but it is not mutually exclusive of antagonistic pleiotropy (Kirkwood, 1997: 685; Crews, 2003: 18). Damage occurs via the gradual accumulation of errors or mutations with age in DNA, which is not transmitted or expressed with exact accuracy or reliability to all of an organism's cells at any age (Finch and Kirkwood, 2000: 65; Lorenzini et al., 2011: 3854).

The **thrifty gene model** (or pleiotrophic gene model) notes that humans, over the course of hominin evolution, became efficient at eliminating plentiful dietary resources (for instance, nutrients; plentiful examples are vitamin C and calcium) and conserving scarce resources (for example, fats, salt, and cholesterol) for early life advantages (Crews, 2003: 43; Myles et al., 2011: 1). Under this model, the genes responsible for regulating bodily control of resources consumed have not changed in modern humans, despite changes in the availability of these resources. As such, these same genes are at the core of a number of risk factors for some chronic degenerative conditions, which are ultimately caused by senescence-related somatic damage – for example, osteoporosis (lack of genetic regulation for calcium conservation), and salt-sensitive hypertension (conservation of salt) (Crews, 2003: 19, 43).

In many developed countries today, the majority of deaths over age 65 are due to heart disease, cancer and stroke (Crews, 2003: 132; see also WHO fact sheet No. 310, 2011; Mathers et al., 2009: 20, 22; Towfighi and Saver, 2011: 2352). Predispositions for these vary according to genes, diet and nutrition and potentially occupation and activities in early life; occupations involving contact with asbestos prior to the realisation that it is dangerous to breathe in is an example. Other chronic degenerative conditions that tend to occur in later life are also affected by the same factors. Cardiovascular disease, coronary heart disease, cancers and diabetes all share (at least some of) the following major risk factors: obesity, fat-rich/fibre-poor diets, tobacco smoke, chemical exposures, nitrates, alcohol consumption, overnutrition, lack of physical activity, hyperinsulinemia, hyperlipidemia, and hypertension (Crews, 2003: 132, 160; Roger et al., 2011; McCracken et al., 2009; Edwards et al., 2010: 547; Chow et al., 2010; Mozaffarian et al., 2009; Bielinski et al., 2012). Exposure to or experience of any combination of these risk factors is largely social and/or cultural – social stress, cultural conformity and lifestyle difference will affect the individual's risk factors (Crews, 2003: 132-133). Socioeconomic status also affects morbidity and mortality – evidence suggests higher morbidity and mortality rates for lower socioeconomic groups, relating to lower educational status, poor housing, low income, or crowding, which lead to poor medical care and poor nutrition (Syme and Berkman, 1976: 1; Grundy and Glaser, 2000; Veugelers et al., 2001; Bassuk et al., 2002; Roger et al., 2011: e38). Psychosocial factors also affect morbidity and mortality, including self-reported health status, social participation and life satisfaction (Schulz-Aellen, 1997: 28-29; McGee et al., 1999; Litwin and Shiovitz-Ezra, 2006; Bowling and Grundy, 2009).

The variation inherent in the individual further enhances differences. For example, while testosterone in males is essential for reproduction, there is some evidence that suggests that men with elevated testosterone levels may have a higher risk of stroke, coronary heart disease and neoplastic diseases of the reproductive organs, as well as correlations with increased risk-taking and anti-social behaviour and other behaviours (Crews, 2003: 110; Stanton et al., 2011; Hyde et al., 2012; Wright et al., 2012). Within the individual there are also gene/environment/behavioural interactions. While it is well known that smoking increases the risk of lung cancer, the disease may occur only in people who possess the genetic predisposition to lung cancer and smoke cigarettes or are exposed to other environmental carcinogens (Crews, 2003: 160-161; Allen, 2012). Similarly, some evidence has suggested the same necessary conditions of genetic predisposition and behaviour in high blood pressure (hypertension) and salt ingestion (Crews, 2003: 139; Gu et al., 2010).

An individual's sex can further influence the types of chronic degenerative conditions to which a person might be susceptible. Osteoporosis is a well known example – low bone mineralisation that worsens over time as bone is resorbed, along with decreasing rates of replacement (Crews, 2003: 145). Women tend to suffer more from this condition than men (Hannan et al., 2000), as men tend to have denser, more robust bones. Changes in the levels of circulating sex hormones result in the decreased rate of bone formation (Glowacki and Zhou, 2007: 81). Osteoporotic bones are less resilient to strain and easier to fracture (Crews, 2003: 145-146). While exercise and diet (usually calcium supplements) act as preventative measures at most ages, they do not add bone mass in those over the age of 65 (Crews, 2003). Clearly, awareness of some conditions and a generally healthy diet before old age may be beneficial in preventing this type of condition in later life.

In some cases, there is a fine line between "disease" and normal progression of ageing, with its inherent accumulation of random somatic damage (Kirkwood, 1997: 690). Osteoarthritis (OA) is one such case, as the most important risk factor is age, although other risk factors include an inherited predisposition, trauma, crystal deposition disease, and, for OA of the knee, obesity (Toivanen et al., 2010; Anderson and Loeser, 2010; Salter and Lee, 2012). However, other factors affect the onset of osteoarthritis, including general "wear and tear", and alterations in matrix components, cell-matrix interactions and chondrocytes (Kirkwood, 1997: 690). While age-related damage accumulation seems to be a major contributing factor, suggesting that OA is at least partly a normal age-related phenomenon (e.g. Weiss and Jurmain, 2007: 445), it is likely that the other factors all play a role, given OA's complex aetiology (Kirkwood, 1997: 691).

Increased understanding of the driving forces of senescence and of "normal" age-related processes, as opposed to pathological processes, can only help uncover the still-poorly understood mechanisms and underlying factors behind the variation in ageing rates. While bioarchaeologists must continue to use the methods at their disposal for estimation of age, it is important that future research seeks to understand such variation in order to develop more accurate and precise age estimation methods, to strengthen our interpretation and understanding of the past.

2.5 Documented Collections of Human Skeletal Remains

The curation of skeletal collections for which age and sex are known for each individual (also called documented collections) provides an invaluable resource for bioarchaeological and forensic anthropological studies. While any skeletal material, archaeological or modern, is important to study, when demographic information is known for the skeletons, different types of questions can be asked. Standards for age and sex determination can be developed and tested using known age and sex skeletal series, but this is not possible with archaeological skeletal samples, because there is no way to test skeletal indicators of age and sex reliably on material for which this information is not known.

Documented collections also provide the opportunity to compare other skeletal traits, reflecting normal and abnormal variation. The study of biomechanics, palaeopathology and growth and development in the past (when non-adult skeletons are also studied) also make use of known age and sex collections. However, bioarchaeology is not the only discipline to benefit from studying documented skeletal collections; palaeoanthropology, neurosurgery and orthopaedics have also used such collections (Tobias, 1991: 278-279).

In the first half of the 19th century, those interested in human bones were trained in anatomy or medicine. Typically, they were more concerned with crania than whole skeletons, particularly to study “racial” differences, and physical/biological anthropology as a discipline did not really come into being until the early 20th century (Armelagos and Van Gerven, 2003: 54; Caspari, 2009: 6; Dias, 1989: 206). Accordingly, one of the earliest American collections was of crania only – Samuel Morton’s collection was intended to educate anatomy students, with curation beginning in Philadelphia, Pennsylvania, USA, around 1830 until his death in 1852 (Buikstra and Gordon, 1981:449). Sir William Turner (1823-1916), of Edinburgh University, is credited with being one of the first to recognise the comparative value of a documented skeletal collection; his collection was in place when Robert J. Terry was a visiting scholar in 1898 (Tobias, 1991: 277).The idea soon spread to anatomy and anthropology departments on the other side of the Atlantic.

Some of the earliest documented collections are still in use today, including the Hamann-Todd Collection (curated at the Cleveland Museum of Natural History, Cleveland, United States; see Meindl et al., 1990), the Robert J. Terry Anatomical Collection (curated at the National Museum of Natural History, Smithsonian Institution, Washington, D.C., United States, and hereafter called the Terry Collection; see Hunt and Albanese, 2005) and the Raymond A. Dart Collection (curated at the University of the Witswatersrand, Johannesburg, South Africa, and hereafter called the Dart Collection; see Dayal et al., 2009); these collections were started at the end of the 19th century by Carl August Hamann (later expanded by Thomas Wingate Todd), Robert J. Terry (inspired after his visit to Edinburgh University,) and Raymond Dart, respectively (Tobias, 1991: 277-278).

Other examples of documented collections include the University of Athens' Human Skeletal Reference Collection in Greece (Eliopoulos et al., 2007), the William M. Bass Donated Skeletal Collection at the University of Tennessee at Knoxville, USA (hereafter called the Bass Collection; Trudell, 1999; Rogers, 1999), the University of Torino's collection of skeletons in Turin, Italy (Giraudi et al., 1984), and the 'A' Series of skeletons of prisoners and poorhouse residents at the University of Helsinki, in Finland (Quigley, 2001:151-152). Usher (2002) published a non-exhaustive list of collections, with a website (www2.potsdam.edu/usherbm/reference; accessed 16/02/2009), including updates; a longer list can be found in Appendix 1.

2.5.1 Challenges

While there is no doubt that documented skeletal collections are enormously valuable to researchers, care must be taken in recognising bias, to avoid overreaching or incorrect interpretation. However, upon recognising bias, increasingly specific research questions can be asked, thus exploiting the same characteristics that may be pitfalls if ignored. For example, it is known that such collections are not representative of the whole populations from which they are derived (Hunt and Albanese, 2005; Ericksen, 1982; Komar and Grivas, 2008; Dayal et al., 2009:10; Tal and Tau, 1983:217; Tobias, 1988:217). In some cases, the goal was to curate specific types of individuals (for example, the pathological collection at the Museum of Pathological Anatomy, in Vienna, Austria – Usher, 2002:40), and thus these collections were not intended to be representative of the population as a whole. Some collections originate from archaeological cemeteries, and will have different characteristics (in terms of representativeness) than collections derived from cadavers originally for teaching anatomy (Albanese, 2003b:2). Other collections, even where an effort has been made to include individuals from all groups in the population at large, generally end up underrepresenting some groups. Many collections, for example, have far more males than females – including the Grant Collection (curated at the University of Toronto, in Toronto, Ontario, Canada), the Maxwell Documented Collection (curated at the Maxwell Museum of Anthropology, University of New Mexico, Albuquerque, New Mexico, USA; Komar and Grivas, 2008: 227), and the Dart Collection (Tal and Tau, 1983: 217). The manner and reasons for curation impact the age and sex structure, the ethnic or cultural groups and the socioeconomic statuses represented in each collection (Albanese, 2003b:2). As Albanese (2003b:3) notes, these differences accordingly impact the types of research questions that can be asked of the collections and the resulting data.

The reasons behind inclusion or non-inclusion of individuals in a documented skeletal collection are similar to those of cemetery collections. Social and cultural biases dictate who is buried in which location (for example, a crypt as opposed to a church yard) and are essentially the same for cadaver-based collections, except that the biases there may be those of the researcher or curator of the collection, deciding which individuals should be included (Albanese, 2003b:22-23). These can be socioeconomic factors, sex and/or gender, ethnic background, religion, age, or occupation. While typically the curator is from the general cultural group whose skeletal remains (s)he is curating (Albanese, 2003b: 22-23), some concepts or ideas of group affiliation may differ between the curator and the individuals belonging to the group. For instance, in South Africa, documented ethnic affiliation for a black person may simply read 'black' or 'South Africa negro', based on identification by a white South African anatomist or doctor – however, the individual, while living, may have self-identified to a particular group (e.g. 'Xhosa' or 'Ndebele'). This may result in bias in the skeletal collection – certain groups may be under- or overrepresented due to differences in the concept of group membership, and thus there may be a lack of detailed documentation.

Albanese (2003b: 18) and Saunders et al. (1995: 110-111) emphasise the importance of recognising the biases inherent in the collection under study; identification of bias is the first step in appropriate use of data from documented collections (whether cemetery or anatomical in origin). Once identified as a bias, erroneous interpretation can be duly avoided – for example, if a collection is found to be skewed towards the lower end of the socioeconomic scale, the researcher would then know not to over-extend interpretation to cover higher status groups. Indeed, these biases can then be exploited in order to ask more specific research questions, controlling for some variables (e.g. socioeconomic status) (Albanese 2003b:21). Documentary and historical records, and their comparison to skeletal evidence, are important tools in recognising bias (Saunders et al., 1995: 110-111).

Excavation, recovery and preservation bias are additional elements adding to the possibility of a cemetery collection being unrepresentative (Walker et al., 1988; Scheidel, 2001a: 11). Small, light bones are easy to miss, resulting in differential recovery (Buikstra and Konigsberg, 1985: 326), particularly when excavators are untrained in archaeology (e.g. cemetery workers may excavate burials for reburial in ossuaries, as in the Coimbra and Lisbon collections), or where there are spatial/site constraints. For example, the Christ Church, Spitalfields excavation (the Christ Church Spitalfields Named Sample, hereafter referred to as the Spitalfields Collection), in London, UK, took place in the church's crypt, and had spatial and safety constraints, although representativeness of the original population may not have been affected (see Cox, 1996: 9). Excavation bias may result when cemeteries are only partially excavated; the possibility always remains that a sub-section of the population was buried in a specific part of the cemetery, or another place altogether, due to age, sex and/or gender, social status or other factors (Scheidel, 2001b: 11). Differential burial practices are known from some societies – for example, on the peninsula between Lake Huron and Lake Ontario, in Ontario, Canada, the indigenous Huron people buried infants in house doorways and under paths often walked by women, so that the souls might be reborn (Penney, 2005: 3, 10). Adults, meanwhile, were buried in ossuaries – although there may have been some exceptions, such as men who died at war (Penney, 2005: 14). Differential preservation can occur with non-adults (or very old adults) as their smaller, more fragile and less calcified bones are more prone to adverse diagenetic effects (Bello et al., 2006: 26-27, 33-34; Walker et al., 1988; Buikstra and Konigsberg, 1985: 326). For church excavation-based collections, people buried in crypts and in lead coffins were usually wealthier (middle and upper class) than those buried in the churchyard (Litten, 2002: 199,225); thus, samples from crypts only are not representative of the population at large, as they consist of a more privileged subset of the population. Burial location within parish churches may also signify different socioeconomic statuses, as intramural burial within the chancel was most sought-after; for those who could not afford intramural vault burial within the church, churchyard space as close as possible to the building was deemed preferable (Litten, 2002: 200,215).

For documented skeletal collections that curate individuals who were first dissected by medical students, differential burial is not an issue, but there is still bias in terms of differences in socioeconomic status. The typical assumption is that bequeathed skeletons tend to be from people higher in socioeconomic status in life, while unclaimed cadavers transferred to anatomy departments tend to be from those who were of lower socioeconomic status in life (Ericksen, 1982: 349; Hunt and Albanese, 2006: 407; Tobias, 1988: 457). However, Patriquin et al. (2002: 105) note that for the Dart and Pretoria Collections (the latter of which is curated at the University of Pretoria, Pretoria, South Africa), the individuals who bequeathed their bodies did so due to inability to pay for burial. Another potential confounding factor is that a person experiencing poverty and low socioeconomic status at death may not have grown up experiencing the same conditions (Ericksen, 1982: 349). For example, Hunt and Albanese (2006: 416) note that while many individuals in the Terry Collection died during the Great Depression and may have been of low socioeconomic status at that time, they may not have experienced low socioeconomic status earlier in life. There is some evidence that changing attitudes towards body donation and subsequent research may be changing the demographic structure of individuals in documented collections; at the University of Tennessee, Knoxville, USA, the Bass Collection's future donors tend to have higher levels of education than current donors, as well as increased diversity (Wilson et al., 2007), implying a change to higher socioeconomic status individuals. They suggest that increased public awareness of body donation for research as an alternative to more traditional burial practices may be the reason.

Another potential issue in the use of documented collections are the reliability of the "known" ages. In cultures or areas where it is uncommon for individuals to know their own exact ages, as in modern non-counting societies, where birth years are not accurately known, or for whom chronological age has no cultural relevance, "age heaping" is frequent (Coale and Demeny, 1983; Scheidel, 1996; Hopkins, 1966). Age heaping is the tendency to report particular terminal digits in stated ages, with the corresponding evasion of other digits; thus, an abundance of ages with terminal digits of 0 and 5 would be evident (Scheidel, 1996; Chamberlain, 2006). This phenomenon can also be found in paleodemographic studies relying on census data, resulting in a nonrepresentative sample population. Age heaping may occur when one individual has reported the age information for all other family members – the individual may only have estimated the family members' ages. For example, in Roman Egypt, it was noticed that 0 and 5 were favoured terminal digits, while the terminal digit 7 was particularly avoided (Scheidel, 1996). The same pattern has been found on mummy labels (labels with the name of the deceased and occasionally other information on Ptolemaic and Roman Period mummies) and tombstone inscriptions. Interestingly, the number 7 was found in many magical spells and charms; Scheidel (1996) suggests that the avoidance of 7 as a terminal digit relates to its ominous magical properties. Avoidance or favouring certain digits is not only an ancient phenomenon – four, for instance, is considered by Chinese people to be unlucky (sounding very much like the word for "death"), while eight is considered lucky as it sounds like the word for "wealth" or "fortune" – many Canadian-Chinese people try to obtain phone numbers with eights in them. While this is simply a anecdotal example, it serves to underline the fact that age heaping may be an inherent bias as a result of the original recorded ages in documented collections, and should be tested for in modern and ancient skeletal samples where documentary evidence of age-at-death is present.

Another potential issue pertaining to paleodemographic census data and reported ages in skeletal collections is that of age inflation of the oldest individuals (Meindl et al., 1983: 73). This may arise when individuals do not know the precise ages of elderly relatives, or it might be to honour their respected oldest family member – a boasting point of sorts. A modern example of age exaggeration is found in Vilcambamba, Ecuador (Mazess and Mathisen, 1982). This population began getting publicity for being extremely long-lived, with many people over 100 years of age. However, subsequent investigation found that age exaggeration was actually at work, with people adding years to their ages from around 60 to 70 years old – and, in fact, no individual was 100 years or older (Mazess and Mathisen, 1982: 518). It has also been suggested that ancient authorities may have had hidden agendas resulting in misleading census data, or perhaps that inefficiency or indifference may have had the same effect (Scheidel, 2001a: 11). Whatever the reason, if such inflations occur, the resulting interpretations of age distributions and demographic reconstructions suffer (perhaps unknowingly) from these artificially inflated distributions (subsequent effects might include incorrect calculations of life expectancy, for example). Accordingly, caution must be taken when extremely old ages are reported in known age skeletal collections.

2.6 Sexual Dimorphism and Sex Determination

This section discusses sexual dimorphism in adult humans and the skeletal sex determination methods that exploit these sex differences. While there are many sex determination methods, both morphological and metrical, the methods that were tested in this thesis are discussed in detail. The morphological methods used here were Phenice’s (1969) pelvic method and Buikstra and Ubelaker (1994) and Walker’s (2008) skull method, alongside Albanese’s (2003a) metrical method.

While many other metrical methods exist (e.g. Purkait and Chandra, 2004, for the femur; Barrier and L’Abbé, 2008, for the radius; Plochocki, 2011, for the sacrum; Wiredu et al., 1999, for the ribs), Albanese’s method was chosen because of the author’s familiarity with it, and because it uses multiple measurements of the femur and pelvis, with different equations allowing for missing data due to fragmentary or damaged skeletons (see section 2.6.6 for more detail). More recently, sex determination methods using geometric morphometrics to assess shape as well as size differences in an objective way have been developed (e.g. for craniofacial traits, see Gonzalez et al., 2011; for mandibular ramus flexure, Oéttle et al., 2005; for the os coxa, Bytheway and Ross, 2010). However, this type of analysis is more expensive than visual assessment of dimorphic traits, requiring specialist hardware and software, and so was not tested in this research.

The morphological methods used here were chosen for their widespread use in bioarchaeology and forensic anthropology and for their high reliability (e.g. Mays and Cox, 2000: 118). The historical developments and early methods of sex determination leading up to the methods tested here are also discussed in chronological order for context.

2.6.1 Introduction and General Trends in Sex Determination

Biological sex, as determined at the moment of conception, is defined by the presence of XX (female) or XY (male) chromosomes. It is the most important parameter to determine for skeletonised individuals in archaeological and forensic contexts, as some skeletal ageing indicators have been found to vary by sex (Brooks and Suchey, 1990; Işcan et al., 1985; Gilbert and McKern, 1973). Thus, sex estimation is required before many of the age indicators can be used.

Sex can generally be assigned with high accuracy (that is, estimated or predicted sex matches actual biological sex); estimates range from 96% allocation accuracy using the pelvis (Phenice, 1969: 300), to accuracy of about 90% with the skull and mandible (St. Hoyme and İşcan, 1989: 69). Sexual dimorphism occurs skeletally during puberty, triggered by the associated hormonal changes, with the development of secondary sex characteristics, especially in the pelvis.

Generally, the skeletons of males are larger and more robust, and females smaller and more gracile, although there is overlap in the intermediate ranges. The pelvis is the most reliable element for sexing the human skeleton, due to necessary structural differences in females associated with childbirth, followed by the cranium (Derry, 1909: 266; MacLaughlin and Bruce, 1990: 1384; Weiss, 1972: 239; Walker, 2005: 385; MacLaughlin and Bruce, 1986: 1380; Meindl et al., 1985b: 85; Bruzek and Murail, 2006: 227). Nearly every other bone has also been used to determine sex (e.g. Steele, 1976; DiBennardo and Taylor, 1982; Steyn and İşcan, 1999; Kemkes-Grottenthaler, 2005; Saunders et al., 2007; Sulzmann et al., 2008).

The expression of sexually dimorphic characteristics is not discretely bimodal (Kelley, 1978: 121), contrary to earlier opinion (Thieme and Schull, 1957: 242), and can vary by age (Walker, 1995; 2005). That is, the cranial and pelvic morphology of younger males tends to appear more "female" than that of older males, while for older females, cranial and pelvic morphology tends to appear more "male"; the cranial morphology of older males may also become more "female" (Meindl et al., 1985b: 81; Walker, 1995: 37, 40; Walker, 2005: 385; Gowland, 2007: 164). Thus, if age is not considered, an age-at-death distribution may overrepresent young females and underrepresent old females; however, the use of the pelvis and skull together can help mitigate these problems (Meindl et al., 1985b: 85). The continuous nature of sexually dimorphic features requires a flexible method including intermediate options for visual assessment systems (e.g. Walker, 2005). There are also metrical methods of sex determination (e.g. Washburn, 1948; Thieme and Schull, 1957; Albanese, 2003a), whose proponents suggest that there is increased objectivity when applying metrical methods and better repeatability, as opposed to visual analysis of dimorphic features (MacLaughlin and Bruce, 1990: 1384). However, others note that some of the anatomical landmarks used in the metrical methods can be difficult to locate, or based on areas that are often damaged in archaeological skeletal material (Thieme and Schull, 1957: 269; Walker, 2005: 385; MacLaughlin and Bruce, 1986: 1381).

In tests of established methods of sex determination, accuracy has often not been as high as that reported in original publications (Lovell, 1989; MacLaughlin and Bruce, 1986; MacLaughlin and Bruce, 1990). Several reasons for the varying levels of accuracy have been suggested. Experience of the observer in the particular method used, or in osteology in general is one such suggestion (MacLaughlin and Bruce, 1990: 1391; Bruzek and Murail, 2006: 227). Variation in the age distribution of the reference sample compared to that of the target sample is another possibility (Lovell, 1989: 119), as expression of some characteristics varies with age (as above, Meindl et al., 1985a; Walker 1995, 2005). Population variation in the expression of sexually dimorphic traits is yet another possibility (Kajanoja, 1966: 32; MacLaughlin and Bruce, 1990: 1391; MacLaughlin and Bruce, 1986b: 230; Walker, 2005: 389; Carlson et al., 2007: 18; Walker, 2008: 48). On the continuum of sexual dimorphism, some populations lay over different sections, with different typical ranges of dimorphism – the sectioning point between typically male morphology and typically female morphology may vary between populations. As such, population-specific standards, often in the form of discriminant function equations for a particular population have been proposed (e.g. Kajanoja, 1966; Birkby, 1966; Steyn and İşcan, 1999; Dayal et al., 2008). Others, however, have found evidence supporting the cross-population use of discriminant function equations (Uytterschaut, 1986: 248-249), although this may be dependent on the populations sampled. Walker (2008: 49) found reliability sufficient in discriminant function equations when used on various modern population samples, but advises caution in applying the same formulae to archaeological samples.

2.6.2 The Pelvis in Sex Determination

Morphological bony pelvic differences between the sexes have long been documented. Turner (1885) wrote about sex classification using the pelvic brim index (using the conjugate and transverse diameters of the pelvic brim), but was more concerned with differentiating between “races” or groups of people, rather than between males and females (Turner, 1885: 127, 131). Derry (1909: 268-272) noted the wide female sciatic notch, relative to the narrow male sciatic notch, and wider female subpubic angles, discussed with regard to childbirth. Others also noted sex differences in the shape and width of the sciatic notch (Straus, 1927: 24-25; Caldwell and Moloy, 1932: 39; Letterman, 1941: 111), and discussed the overlap in expression of sexually dimorphic pelvic traits – individuals could display a mix of “male” and “female” features and measurements (Straus, 1927: 27; Letterman, 1941: 115-116) – and that females with higher degrees of “male” morphology tended to run into obstetrical problems (Caldwell and Moloy, 1932: 39).

2.6.3 The Ischiopubic Index

Washburn (1948) presented the first known formal metrical sex discrimination method tested on a known sex sample, where pubic and ischial lengths were measured to calculate an index. The method was developed on 300 known sex individuals from the collection at Western Reserve University, USA (presumably the collection that is now called the Hamann-Todd Collection) (Washburn, 1948: 200). The measurement landmarks were later criticised as being difficult to locate (Stewart, 1954: 390; Thieme and Schull, 1957: 269; MacLaughlin and Bruce, 1986: 1381). The index was found to accurately sex 90% of individuals (higher if used with sciatic notch morphology), with the caveat that 'races' be treated separately, as differences were found between black and white individuals (Washburn, 1948: 206). This method was later tested on Bantu and Bushman individuals from the Dart Collection (Washburn, 1949: 428-429) and on an Alaskan Eskimo sample (Hanna and Washburn, 1953: 25), with allocation accuracies of 98% and 95%, respectively, but with some overlap between modes (sexes). Stewart (1954: 389-390), however, believed that important morphological information was lost by using linear measurements, and the main application should be for skeletons whose sex was uncertain, with intermediate morphology.

2.6.4 The Phenice Method

In 1969, Phenice published a visual analytical method of determining sex by scoring three morphological traits, essentially for presence or absence, which was tested on a sample from the Terry Collection. The three traits are subpubic concavity, ventral arc, and the medial aspect of the ischiopubic ramus (Phenice, 1969: 298, 300, see Figure 2.1, below). In the typical female, the ventral arc is present, as 'a slightly elevated ridge of bone which extends inferiorly across the ventral surface to the lateral most extension of the subpubic concavity'; if a slight elevation is present in a male, the course of the ridge extends in a more inferior direction, parallel to the medial border, or infero-medially (Phenice, 1969: 298, 300). The ventral arc is the origin of the gracilis muscle; the position of the origins of this muscle (as well as that of the adductors magnus and brevis, but to a lesser degree) differs in males and females, corresponding with the varying course of the ventral arc in females compared to the slight ridge (or its absence) in males (Anderson, 1990: 453). The medial elongation of the pubis in females, especially at the pubic symphyseal inferior margin (Coleman, 1969: 141) and commencing at puberty, accounts for the differential muscle position in males and females (Anderson, 1990: 454; Budinoff and Tague, 1990: 78). As growth occurs in pubertal females, the pubic symphysis (and joint capsule) begins to encroach laterally onto the ventral surface, 'pushing' the muscles in a more lateral and inferior direction; as such, males are not likely to show a similar ventral ridge course (Anderson, 1990: 457). The ventral arc was the most reliable trait for sex assessment in Phenice's sample. While male subpubic concavity is convex to slightly concave, female subpubic concavity is concave. The medial aspect of the ischio-pubic ramus in females tends to be a narrow ridge, while in males, this area tends to be flat and broad (Phenice, 1969: 300). While Phenice (1969: 300) noted that not all individuals will display all female or all male traits; however, Phenice felt that there was 'almost always' at least one trait clearly signifying whether the individual was male or female. Accuracy ranged from 94.44% (for black pelves) to 96.55% (for white pelves) (Phenice, 1969: 298).

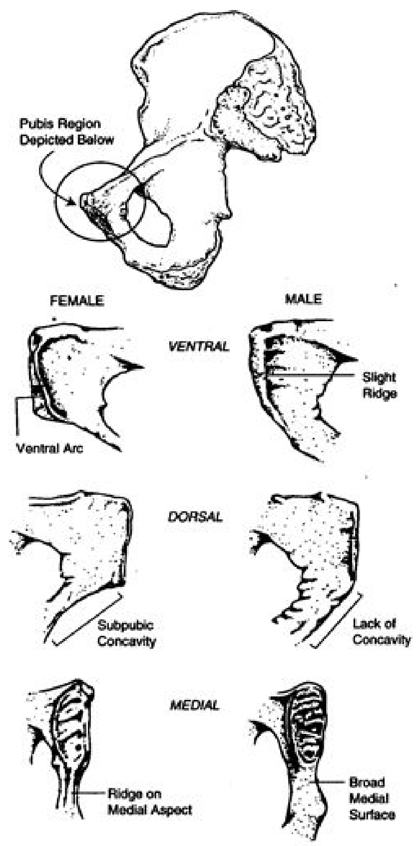


Figure 2.1. The pubis region and sex differences used in the Phenice method: the ventral arc, subpubic concavity and ischiopubic ramus ridge. Right side. From Figure 1, Buikstra and Ubelaker, 1994: 17.

Subsequent tests of the Phenice method (on documented and archaeological samples) have not had quite as much success in accuracy, and have noted the need to score intermediate morphologies (Lovell, 1989: 118; Kelley, 1978: 121; MacLaughlin and Bruce, 1990: 1389). Interestingly, in Kelley's (1978: 121) archaeological sample of California Indians, intermediate morphologies were found more often in females, perhaps suggesting increased variability in female pubic morphology than male. Population variation was another suggested reason for the lower accuracy (MacLaughlin and Bruce, 1990: 1391).

2.6.5 The Sciatic Notch

As mentioned earlier, the sciatic notch is known to show sexually dimorphic variation in shape and size; the male sciatic notch tends to be U-shaped and narrow, while the wider female sciatic notch tends to be relatively more shallow (Walker, 2005; Bruzek, 2002; Singh and Potturi, 1978). In attempts to make sex assessment from the sciatic notch more objective, some researchers have tried to quantify descriptions of this variation (Letterman, 1941; Singh and Potturi, 1978; DiBennardo and Taylor, 1983; MacLaughlin and Bruce, 1986b; Milne, 1990), although visual assessment of morphological differences seems to be more widely used in evaluating sex using the sciatic notch, possibly because landmarks used in metrical methods are often damaged or lost in archaeological material or difficult to locate (Walker, 2005: 385). Walker (2005: 386) proposed an ordinal scoring scale, changing slightly Acsádi and Nemeskéri‘s (1970: 84) earlier scale that went from -2 to +2, because he found the androgynous 0 sometimes difficult to apply to other populations where that particular morphology was not an appropriate point of discrimination between male and female. The ordinal scale (also in Buikstra and Ubelaker, 1994; see Figure 2.2, below) is from 1 to 5, where 1 is the most female morphology and 5 is the most male morphology. Samples from the Hamann-Todd Collection, the Terry Collection and the St. Bride’s Church collection (a crypt sample from London, UK) were used to test the scoring method and variation between populations in notch morphology (Walker, 2005: 386). Although a surprising number of individuals with a score of 2 were male, 88% of individuals with a score of 1 were female, and 91% of individuals with a score over 2 were male, supporting the use of sciatic notch morphology as a sex discriminator (Walker, 2005: 389).

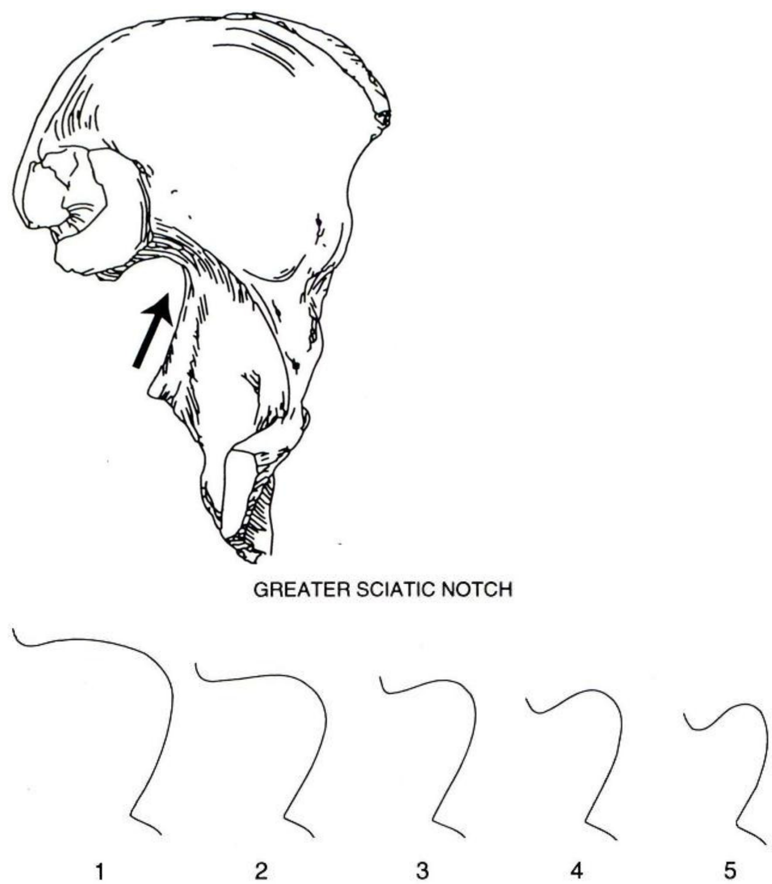


Figure 2.2. Left os coxa with an arrow indicating the sciatic notch and schematic representation of the sciatic notch with possible scores for sex determination. From Figure 2, Buikstra and Ubelaker, 1994: 18.

2.6.6 Albanese's Metrical Method

A metrical method was presented by Albanese (2003a), and tested in this research. As well as using two pelvic measurements, this method also uses several measurements of the femur, which has sexually dimorphic characteristics and dimensions in various populations (Parsons, 1914, 1915; Van Gerven, 1972; Purkait, 2005). Albanese's (2003a: 4) pelvic measurements use a new acetabular landmark – the supero-anterior apex of the lunate surface, purported to be easier to identify than landmarks used in earlier studies. The measurements used in this multivariate method are maximum femur length, maximum femoral head diameter, epicondylar breadth of the femur, hipbone height, iliac breadth, the superior pubis ramus length (SPRL) and acetabular-ischium length (AIL). Both SPRL and AIL use the acetabulum's supero-anterior apex of the lunate surface; SPRL is the maximum length from that landmark to the pubic symphyseal superior margin, while the AIL is the maximum length to the most inferior point on the ischium (Albanese, 2003a: 4). Logistic regression was used to develop formulae with various combinations of the measurements, so that measurements can be input to allocate sex (Albanese, 2003a: 8). The reference sample used to develop this method consisted of individuals from the Terry Collection and individuals from the Coimbra Collection (Albanese, 2003a: 2). If the resulting probability is less than 0.5, assigned sex is female; if the probability is more than 0.5, assigned sex is male. Accuracy was 98.5% on a test sample for males and females (Albanese, 2003a: 7).

2.6.7 The Skull in Sex Determination

As mentioned above, the cranium is considered the second-most reliable indicator of sex in humans (Derry, 1909: 266; MacLaughlin and Bruce, 1990: 1384; Weiss, 1972: 239; Walker, 2005: 385; MacLaughlin and Bruce, 1986: 1380; Meindl et al., 1985b: 85; Bruzek and Murail, 2006: 227). Sex differences were noted in the skull fairly early (Parsons and Keene, 1919; later, Keen, 1950), and such differences were often measured and quantified. Size differences were emphasized; the larger, heavier skulls of males, with heavier muscle markings, were contrasted with the more 'infantile' type female skull (Keen, 1950: 66). The supraorbital ridges and nuchal lines on the occipital bone were scored as absent, medium or marked (Keen, 1950: 70). Keen (1950: 75) noted that sex allocation could be up to 85% accurate with the skull alone; however, neither non-adult or skulls of the elderly were appropriate for sex discrimination, as sex differences did not appear until after puberty, and 'because senile changes tend to disturb the sexual expression of many skull features'.

Metrical methods and their subsequent testing followed (Giles and Elliot, 1963; Kajanoja, 1966; Birkby, 1966). Tests of populations different in geographic origin compared to the reference collection reported lower allocation accuracies, and population-specific discriminant function equations were suggested (Kajanoja, 1966: 32; Birkby, 1966: 26). Discriminant functions based on the crania of specific populations, using linear measurements or geometric morphometric methods are still being developed (Cunha and van Vark, 1991; Steyn and İşcan, 1998; Franklin et al., 2005; Dayal et al., 2008).

The advantages of visual assessment over metrical methods are speed of observation and assessment, that shape can be evaluated (as well as size), and that it is inexpensive, as no equipment is necessary (Walrath et al., 2004: 132-133; Walker, 2008: 39); of course, intra- and interobserver error exists for both visual and metrical methods of sex determination. Further, as Walrath et al. (2004: 133) note, as visual assessment is subjective, adjustments can be made in allocation to account for interpopulation variation. Acsádi and Nemeskéri (1970: 88) seem to have presented the first formal scoring system for these visually assessed traits, with schematic diagrams for scoring the glabella, nuchal crest, mastoid process, supraorbital margins, mental eminence, and mandibular angle. These diagrams and descriptions were updated and clarified in Buikstra and Ubelaker's *Standards for Data Collection from Human Skeletal Remains* (1994, see Figure 2.3, below), the result of collaboration between many biological anthropologists, in order to standardise data collection to improve comparability of data between studies. The method involves scoring the nuchal crest, mastoid process, supraorbital margin, glabella (or supraorbital ridge), and mental eminence, with a scoring scale ranging from 1 to 5 (Buikstra and Ubelaker, 1994: 20), modified from Acsádi and Nemeskéri's (1970: 87, 89) scale of -2 to +2. Walker (2008: 40) produced the diagrams used in *Standards* and tested his modifications on samples from the Hamann-Todd and Terry collections, the St. Bride's Collection and an archaeological sample of Californian Native Americans. Logistic discriminant analysis models were produced from the scores, with allocation accuracy ranging from 84 to 88% (Walker, 2008: 49). While intraobserver error was fairly low, interobserver error was sometimes statistically significant (particularly for traits without accompanying diagrams), meaning that comparisons of data collected by different observers should be attempted only with caution (Walrath et al., 2004: 136). Inter- and intra-population differences were found (Walker, 2008: 48). As such, caution is advised when using discriminant function equations developed on modern populations on archaeological samples. Other studies have presented sex discrimination methods using the same variables as presented by Walker (2008), but with different statistics to allocate sex (Konigsberg and Hens, 1998; Stevenson et al., 2009), or with more variables to assess (Williams and Rogers, 2006), with varying levels of allocation accuracy.

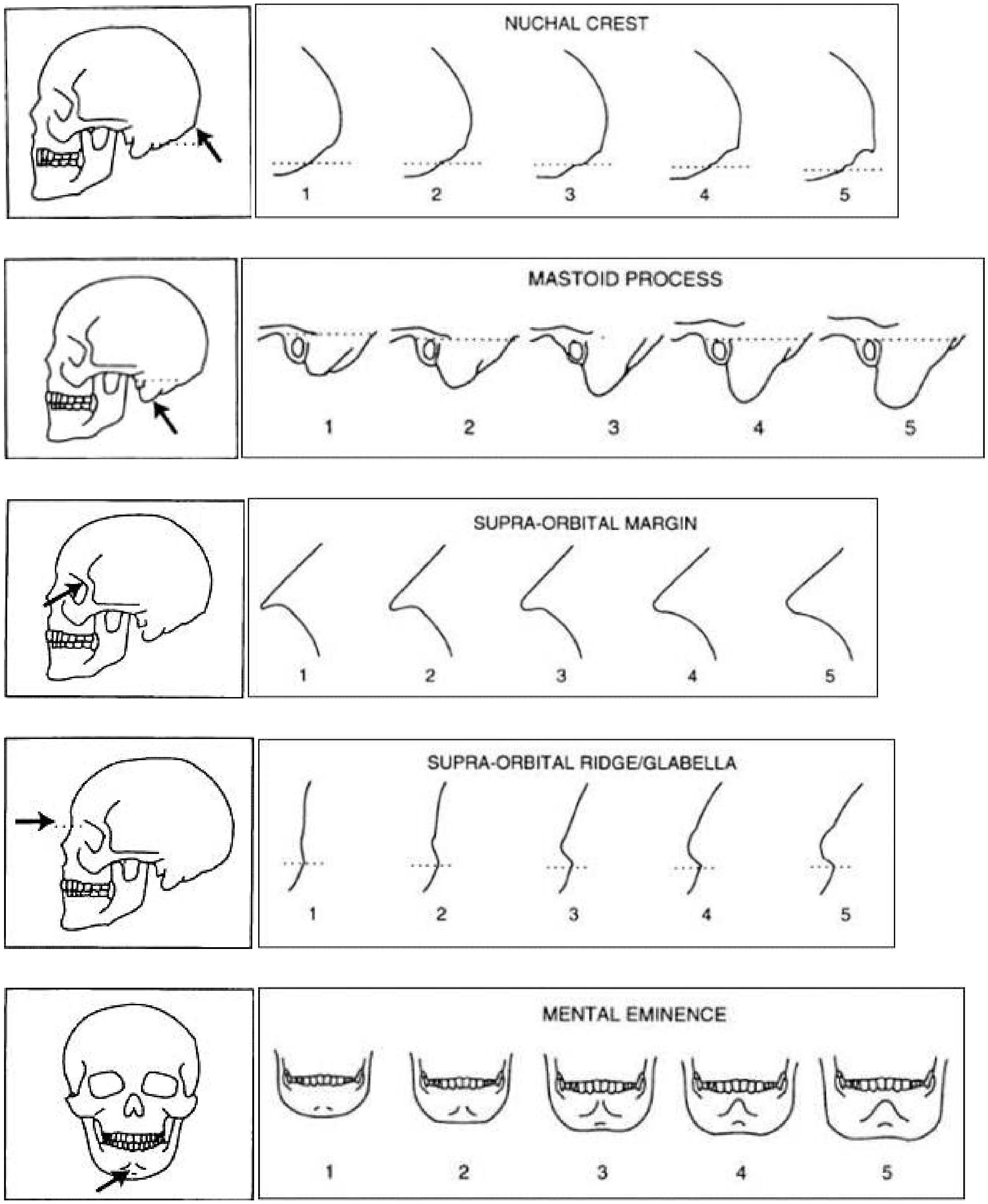


Figure 2.3. The skull with scored traits used for sex determination: nuchal crest, mastoid process, supraorbital margin, glabella and mental eminence. From Figure 4, Buikstra and Ubelaker, 1994: 20.

2.7 Age Estimation and Differential Ageing Rates

2.7.1 Introduction and General Trends in Age Estimation

Estimating age-at-death is essential for reconstructing paleodemographic profiles, as has been discussed. In non-adult skeletons, estimating age is a relatively straightforward process. This is because it relies on analysis of growth and development, biological processes that are more regularly scheduled, generally resulting in smaller standard errors (see Scheuer and Black, 2000; Lewis, 2007; Gowland, 2007). While there is evidence for variation in growth and development schedules between populations, ageing of adults is more highly variable as it relies on recording and interpretation of degenerative changes to skeletal elements, particularly of those joints believed to be minimally influenced by biomechanical processes (e.g. the pubic symphysis, cranial sutures) (Chamberlain, 2006: 105). However, there is considerable variation in the rates of degeneration both between populations and between individuals within populations, as it is perhaps more dependent on environmental, cultural and behavioural factors than is growth and development (e.g. Kemkes-Grottenthaler, 2002; Chamberlain, 2006: 105; Falys and Lewis, 2011: 705).

Furthermore, there is no particular reason why degeneration, an indication of biological age, should correspond with chronological age, as chronological age is simply one method of counting one's lifespan, and one calendar year does not have any specific biological significance. As the same range of factors affect individual rates of ageing, perhaps with variable contributions by individual, biological age varies by individual for chronological age (Karasik et al., 2005: 575). Contributions of various factors may also vary over the life course. For instance, activity level and intensity change with age, suggesting that activity's contribution may slow with age in individuals whose activity lessens (Karasik et al., 2005: 578).

Growth and development are tied to life stages (e.g. Todd, 1921d: 336), for example, infancy or adolescence; after maturity is achieved, and reproduction can begin, environmental and cultural factors have more influence than in earlier life stages and more variation is seen in the ageing process (Wittwer-Backofen et al., 2008: 390). As such, the ageing of adults is more problematic, and error and imprecision increase with age (Buikstra and Konigsberg, 1985; Meindl and Russell, 1998; Schmitt, 2004: 2; Bedford et al., 1993; Katz and Suchey, 1986; Murray and Murray, 1990; Nagar and Hershkovitz, 2004: 153; San Millán et al., 2013: 1749). To mitigate this problem, the consensus among biological anthropologists is that multiple ageing methods must be used for the best possible age estimate (Meindl and Russell, 1998, Van Gerven and Armelagos, 1983). Indeed, many researchers agree that younger adult individuals are typically overaged, and older individuals are typically underaged (Bedford et al., 1993: 290; Cox, 2000; Kvaal et al., 1994: 365, 367). This is the so-called 'attraction of the middle' (Masset, 1989: 82), or a result of the regression techniques used in developing the methods of age estimation for age ranges and means for predicting age imposing the structure of the reference sample on the target sample (the sample being assessed for age-at-death of its individuals) (Nagar and Hershkovitz, 2004: 151; Aykroyd et al., 1999: 61; Hartnett, 2010a: 1149-1150). That is, individuals with the oldest actual ages in the reference sample will cause the regression line to rise, increasing the slope of the line (Hartnett, 2010a: 1150).

Variability in ageing in females has been found to be wider than in males (Jackes, 1985: 291; Brooks and Suchey, 1990: 233; Djurić et al., 2007: 22), although Hartnett (2010a: 1149) found the opposite – that females seemed to have a 'more regular and predictable pattern' in pubic symphyseal morphological change. This can result in wider age ranges for female age estimates, and potentially lower accuracy (Jackes, 1985: 291).

Most age indicators have fairly wide associated age ranges, which are necessary to encompass at least a portion of normal variation in the ageing process (Hartnett, 2010: 1150). The age ranges associated with various age indicators vary in width; even if the standard is to predict age within ± 2 standard deviations or 95% confidence intervals (CI), this means that it is difficult to compare accuracy using percentages of correctly aged individuals, as the range of 95% CI might be significantly wider for a particular ageing method compared to another (Klepinger et al., 1992: 768). A method with wider intervals will "accurately" predict the ages of more individuals than a method with narrower intervals. A related phenomenon is that a method with wider intervals likely also has fewer potential age categories or phases, making it more likely that an age will be "accurately" predicted than a method with more age categories or phases (and narrower age ranges).

There are at least as many tests of age determination methods as there are methods. Practically without fail, any test of a method where the testee was not a developer of the method finds that the method does not perform as well as claimed by the developer of the method (e.g. Suchey, 1979; Murray and Murray, 1991; Saunders et al., 1992). While interobserver error and/or inexperience with the method may be to blame, another likely culprit is human variation (e.g. Katz and Suchey, 1989; Hoppa, 2000; Schmitt, 2004). As a result, many researchers have called for population-specific standards (Schmitt, 2004; Oettlé and Steyn, 2000; Hoppa, 2000; Kemkes-Grottenthaler, 2002: 65; Baccino and Schmitt, 2006; Kimmerle et al., 2008: 530). This refers to a calibration of an ageing method suitable for a particular population, with separate ranges of values for methods of age and sex determination that are specific to a certain geographical location and/or time period. While this would be ideal, it is probably an impossible goal – known age and sex collections are necessary for calibration, and these are not available for every population globally, and certainly not for archaeological populations (İşcan and Loth, 1989).

The reasons for these differences in ageing rates for different skeletal indicators are complex, and will be discussed further in the Human Variation section.

2.7.2 Age Estimation Methods and Other Considerations

Kemkes-Grottenthaler (2002: 53, based on Spirduso, 1995: 47) has outlined primary, secondary, and tertiary criteria that a 'gerontological biomarker' (age indicator) must meet:

Primary criteria:

- Strong correlation between biological feature and age
- Age indicator is not altered by pathological events
- Age-related alteration is not secondary to metabolic or nutritional changes
- Sequential and unambiguously identifiable ageing pattern
- Continuous remodelling throughout lifespan

Secondary criteria:

- Wide applicability
- Generalization across species

Tertiary criteria:

- Reliable changes within a short time interval as compared with total lifespan

(Kemkes-Grottenthaler, 2002: 53)

No single skeletal age indicator fulfills all of these criteria; thus, Kemkes-Grottenthaler (2002: 53) concludes that all skeletal age indicators are 'inherently flawed'. Another major criticism of age estimations in general came from Bocquet-Appel and Masset (1982), when they stated that the age distribution of the target sample mimics that of the reference sample, and some other studies have provided further evidence of this phenomenon (e.g. Aiello and Molleson, 1993: 702). While they were discussing this problem in the context of paleodemography and age-at-death distributions, it should be noted that it also affects individual age estimates. The use of multiple methods, outlined above, is one way to attempt to correct these flaws.

2.7.2.1 Statistical Methods

Bayesian statistics to estimate age distributions has also been recommended (Konigsberg and Frankenberg, 1992; Hoppa and Vaupel, 2002; Chamberlain, 2000, 2006; Steadman et al., 2006). This approach uses the raw age indicator data in terms of phases (or scores, depending on the method) that the researcher has assigned to each individual and prior probabilities of that indicator phase belonging to particular age categories, to solve for the conditional probability of the individual being in a particular age category, given that they exhibit that age indicator phase. The prior probabilities are derived from maximum likelihood estimation of age distribution on the target sample indicator states, while the probability of the indicator phase belonging to a particular age category given the age category is taken from the reference sample (Chamberlain, 2006: 112-114). Other options for selection of prior probabilities include assuming a uniform prior probability, or using model life table mortality rates (Chamberlain, 2006: 114-115). The best results from the use of Bayesian methods seem to be from mortality model-constrained maximum likelihood estimations of target sample age distributions; otherwise, unrealistic age distributions may result. Transition analysis is another statistical method used to generate age distributions and estimations, but this examines the timing of the morphological change from one phase to another (the 'transition') (Boldsen et al., 2002; Konigsberg and Hermann, 2002).

2.7.2.2 Other Age Estimation Methods

While the statistical methods may certainly prove valuable after further testing, the complexity of performing such operations seems to have made many researchers shy away from their use in every-day skeletal analysis. Indeed, tests of methods of age estimation using morphological skeletal change of various elements are still forthcoming. Other categories of age indicators include:

- microscopic methods, such as osteon counts and percentages of unremodelled bone (see Kerley, 1965; Bouvier and Ubelaker, 1977; Keough et al., 2009; Cannet et al., 2011; Castillo et al., 2012);
- macroscopic dental methods (dental root dentine translucency: e.g. Bang and Ramm, 1970; Beyer-Olsen et al., 1994; Schmitt et al., 2010; calculation of pulp-tooth volume ratios: e.g. Star et al., 2011)
- microscopic dental methods (such as cementum annulation counts: see Stott et al., 1982; Charles et al., 1986; Joshi et al., 2010; Gauthier and Schutkowski, 2013);
- dental wear (e.g. Gustafson, 1950; Murphy, 1959; Brothwell, 1981: 71; Scott, 1979; Mays, 2002; Gilmore and Grote, 2012);
- radiography to assess bone loss (Todd, 1930; Acsádi and Nemeskéri, 1970; Walker and Lovejoy, 1985; Wade et al., 2011; Curate et al., in press);
- biochemical methods (for example, amino acid racemisation of dental enamel: e.g. Masters and Zimmerman, 1978; Griffin et al., 2009, for a test of this method; Sakuma et al., 2012).

While these are all certainly useful in various contexts, for example, as part of a multivariate estimate, or if the more "traditional" ageing elements are not well preserved in a given skeleton, each also has disadvantages. The microscopic and biochemical techniques may involve destruction of the bone (which is not possible for some researchers or in some countries where skeletal collections are curated), and/or can be expensive to use (and require expertise or training); dental wear is notoriously population specific, as the grittiness of diet plays a major role in the rate of tooth wear (Molnar, 1971). Finally, some of these methods require quite elaborate sample preparation that might not be feasible for some projects (for example, some researchers do fieldwork in other countries, and are not able to transport samples easily to home institutes for such analysis).

Morphological skeletal age indicators remain among the most well-used methods of age estimation; advantages include their non-destructive nature, relative ease of application and that they are inexpensive and do not require specialist equipment. In general, the methods requiring visual assessment are most commonly used in forensic anthropology; these are the pubic symphysis, auricular surface, sternal rib ends and cranial suture closure (Garvin and Passalacqua, 2012: 3). For these skeletal regions (respectively), the Suchey-Brooks (Suchey, 1979; Katz and Suchey, 1986; Brooks and Suchey, 1990), Lovejoy et al. (1985b) (and followed by Buckberry and Chamberlain, 2002), İşcan et al. (1984a), and Meindl and Lovejoy (1985) methods were the most commonly employed (Garvin and Passalacqua, 2012: 3). Because these are the most commonly used methods, they were chosen for testing in this research; detailed descriptions are provided in the following sections, along with the prior related historical developments and early methods.

2.7.3 Cranial Suture Closure

The reasons for cranial suture closure, though seemingly age-related, are not completely understood. If age were the only factor, a much higher frequency of complete closure at the older ages would probably be expected. However, it is fairly common for old individuals to have open or partially open sutures (e.g. Key et al., 1994; Powers, 1960). In infants, open sutures allow for neurocranial expansion to accommodate brain growth; the brain signals to the cranium to expand via the dura mater (Opperman, 2000: 481). These signals stimulate new bone to be formed at the edges of the calvarial bones, but to remain unossified, with undifferentiated cells within the suture to allow for continued later growth, and maintain the approximate width of the sutures themselves. In this way, cranial sutures function as intramembranous bone growth sites (Opperman, 2000: 481). It is possible that the undifferentiated cells within the sutures are recruited from the dividing cells in the osteogenic cell layer that line bone fronts. When sutures develop, cells in this layer can divide, differentiate to begin producing bone-related proteins, or become apoptotic (apoptosis is the process of programmed cell death). Some cells in the suture matrix also become apoptotic – it is assumed that this is to prevent an increase in cells in the suture matrix (Opperman, 2000: 482). Disruption or imbalance in these functions may cause craniosynostosis (the premature closure of cranial sutures, Nieminen et al., 2011: 67; FitzPatrick, 2013: 231). Opperman (2000: 482) notes that in examination of the factors necessary to maintain open sutures, there is some difficulty in understanding because much is still unknown about cell expression and regulation factors ‘during “normal” suture morphogenesis and growth and what constitutes sufficiently altered levels to result in suture obliteration.’ This implies that there is also a lack of knowledge of the “normal” pattern or rate of suture morphogenesis, and whether regulation of cell functions on the bone fronts and within the matrix are perhaps less important when the brain is no longer growing.

The role of biomechanics in maintaining suture patency (that is, sutures remaining open) is also unclear (Opperman, 2000: 482). Other factors that have been implicated in suture closure include genetic, hormonal, vascular and local factors (Cohen, 1993: 593). This, alongside the lack of clarity on whether suture closure is indeed the ‘natural’ end result of growth (Cohen, 1993; 593), makes it more difficult in appraising the role of suture closure in age estimation. It seems that the termination of growth does not definitively mean that sutures will close, as cranial sutures remain open for at least some time after the brain ceases growth; the circummaxillary suture also remains open until fairly old age (the seventh and eighth decades), but facial growth finishes by about 20 years (Cohen, 1993: 593-594). Indeed, Cohen (1993: 594) notes that ‘Patency is the original condition evolutionarily and ontogenetically; it is fusion that needs to be explained.’

Cohen (1993: 590) notes that, in terms of suture morphology, the longer a suture remains open before obliteration, the more complex it becomes, in the form of serration and interdigitations. These interdigitations may help resist mechanical stress, from shearing, compression and/or tensile forces. Hydrocephalus (‘enlargement of the normal fluid-containing spaces within the brain substance, associated with increased pressure due to accumulation of fluid’, Roberts and Manchester, 2007: 53; Munch et al., 2012: 2409; Zak et al., 2012: 186), if left untreated over a long period of time, results in profound interdigitations, providing evidence for the effect of force on suture structure (Cohen, 1993: 590).

2.7.3.1 History of Study

While ancient scholars recognised that cranial sutures were present and closure varied by individual (for example, Hippocrates, *On Injuries of the Head,* writing around 400 BCE), the investigation of the relationship between cranial suture closure and age began in the latter half of the 19th century (Pommerol, 1869; Sauvage, 1870; Dwight, 1894). While the age at which suture closure begins was variously found to be between 30 and 45, several studies agreed that endocranial closure preceded ectocranial suture closure (Pommerol, 1869:20; Sauvage, 1870: 582-583; Dwight, 1894: 97; Parsons and Box, 1905: 32, 34, 37). Parsons and Box (1905: 37) found that endocranial sutures were more reliable for age estimation. Sex differences were also noted:

male ectocranial sutures were found to begin closure earlier and be obliterated earlier than those of females (Parsons and Box, 1905: 32, 38).

2.7.3.2 Todd and Lyon

Todd and Lyon (1924) were the first to publish a large systematic study of suture closure and its relationship with age, developing a scale of closure (0 to 4, from open to complete closure) and specific age ranges at which changes occurred, including beginning of closure and completion. Both endocranial and ectocranial suture closure in white and black adult males were analysed (Todd and Lyon, 1924; Todd and Lyon, 1925; Todd and Lyon, 1925b; Todd and Lyon, 1925c); a total of 514 skulls from Western Reserve University's collection were used for these studies, including small numbers of white and black females. Sutures were grouped into three categories: vault, circum-meatal, and the accessory sutures (Todd and Lyon, 1924: 336). No significant differences in closure pattern between the white and black adult males, or between males and females were found (Todd and Lyon, 1924: 333; Todd and Lyon, 1925b: 48). Contrary to previous research, they also found that endocranial sutures did not necessarily begin closure before ectocranial sutures, but that ectocranial sutures tended to have a slower and more variable closure rate, suggesting that this may have accounted for previous findings (Todd and Lyon, 1925: 23-24; Todd and Lyon, 1925c: 151).

The main criticisms of Todd and Lyon's work are of the veracity of the "known" ages and the rejection of large numbers of "abnormal" crania (Meindl and Lovejoy, 1985: 57-58). The known ages of the skeletons at Western Reserve University's collection were based on city records and hospital records, although Todd (1920: 289) noted that not all people know their age, or misstate it for a variety of reasons. The records themselves were found to be disorganised and some untrustworthy: '…the data which we were gathering in the laboratory were far more trustworthy than official documents' (Todd, 1920: 289). To exclude the skeletons with uncertain ages, the skeletons' "known" ages were checked against the skeletal age indicators, so that those with wide discrepancies between known and skeletal age were rejected (Todd, 1920: 289-291); however, this results in somewhat circular reasoning. It was noted, after excluding the rejected crania, that suture closure times now seemed 'remarkably uniform' (Todd and Lyon, 1924: 341).

The next wave of literature on cranial suture closure seems mostly critical of Todd and Lyon's work, primarily in that using their data on suture closure to estimate age on other skeletons did not seem to work very well (McKern and Stewart, 1957; Powers, 1962; Cobb, 1955; Genoves, 1960). McKern and Stewart (1957: 27, 37), in their work on Americans who were killed or died as prisoners of war (POWs) during the Korean War, noted that for age-at-death, suture closure can be used only for 'crude estimates in terms of decades only'. Others agreed (Ashley-

Montagu, 1938: 372; Cobb, 1955: 394; Brooks, 1955: 577; Powers, 1962: 54), and suggested their use only as part of a multivariate age estimate (Acsádi and Nemeskéri, 1970: 120), or recommended abandoning their use altogether (Genoves, 1960: 207). Powers (1962: 54) found that cranial suture closure overestimated ages for those aged between 25 and 35 years, and underestimated age in those over 60 years with open sutures; the best results were for individuals between 20 and 28 years, and individuals over 68 years were not aged reliably.

2.7.3.3 Meindl and Lovejoy

Meindl and Lovejoy (1985) presented a new method of age estimation using cranial suture closure, after a period in which cranial sutures seemed to have been abandoned as an age indicator (in the literature, if not in practice). A sample from the Hamann-Todd Collection was used, consisting of 236 crania, chosen for the reliability of known age. In developing this method, the ectocranial sutures alone were used, as these are purported to close at older ages, and a method to age older individuals, particularly for forensic contexts, was desired (Meindl and Lovejoy, 1985: 58). While 17 ectocranial suture sites were preliminarily analysed, and scored from 0 to 3 (from open to closed; see Figure 2.4 for examples), the suture sites offering the best age-related information were grouped into two sets: the vault system, and the lateral-anterior system

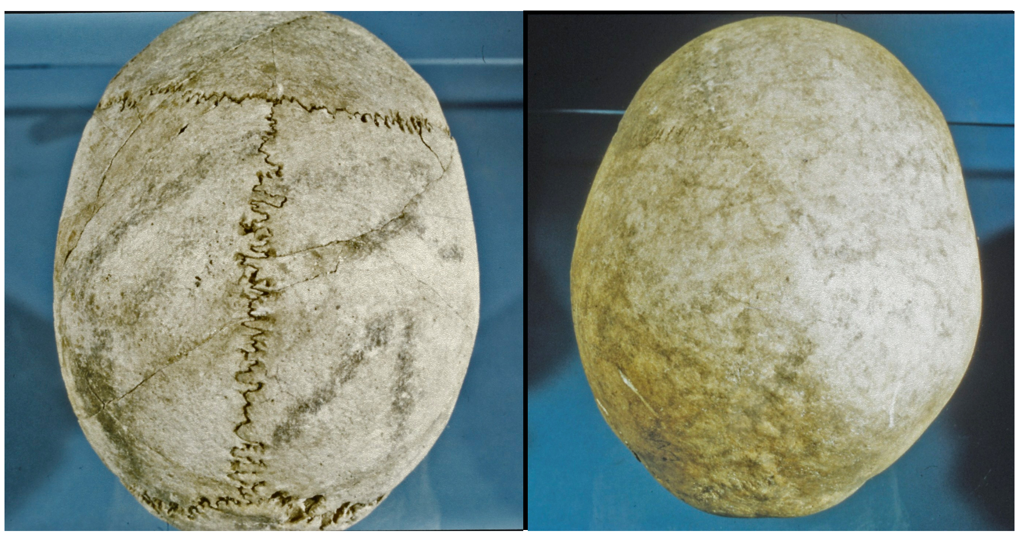


Figure 2.4. Examples of open and closed cranial sutures. Images courtesy of Prof. C. Roberts. Using Meindl and Lovejoy's (1985) method, the skull on the left would have a score of 0 (open sutures), and the skull on the right would have a score of 3 (closed sutures). The frontal bone is towards the top of the photographs.

(Meindl and Lovejoy, 1985: 60). The vault system sites comprise the midlambdoid, lambda, obelion, anterior sagittal, bregma, midcoronal, and pterion; the lateral-anterior system consists of the sphenofrontal, inferior sphenotemporal, and superior sphenotemporal alongside the pterion and midcoronal. The scores for each suture site are added into a composite score for each system, and then given an age estimate based on this score. Ranges are wide, and standard deviations range from 6.2 years to 12.6 years. The lateral-anterior system was found to be more useful for older ages, but for crania with completely closed sutures, it was recommended that other ageing methods should be used instead (Meindl and Lovejoy, 1985: 62). Similarly, for crania with completely open sutures, it was recommended that postcranial indicators should be used to estimate age instead of the sutures.

It was noted that the age and suture closure relationship is 'only general', as there is much overlap of age ranges and variability (Meindl and Lovejoy, 1985: 62). An analysis of variance was undertaken to look for sex or "race" differences, as previous studies were not in agreement on this point (e.g. Brooks, 1955; Todd and Lyon, 1924; 1925); Meindl and Lovejoy found '*no measurable bias to age prediction'* due to sex or 'race' (1985: 64, italics theirs) using either the lateral-anterior or vault system. The authors conclude by warning again against using any single indicator to estimate age, as no sole indicator 'is ever likely to accurately reflect the many factors which accumulate with chronological age', but suggesting that their method may help to add value to multivariate age estimates (Meindl and Lovejoy, 1985: 65-66).

2.7.3.4 Tests of Meindl and Lovejoy

Key et al. (1994) later compared Meindl and Lovejoy's (1985) method to the Acsádi-Nemeskéri endocranial closure method on a sample of 183 individuals from the Spitalfields Collection. The Acsádi-Nemeskéri method was found to sort skulls into general age categories only – young, middle-aged and old – and only up until age 50, as little endocranial closure activity occurs after this age (Key et al., 1994: 197, 206). When the Meindl-Lovejoy method was used, sizeable differences in mean age per stage were found between the Spitalfields sample and those given by Meindl and Lovejoy, based on the Hamann-Todd Collection (Key et al., 1994: 200). The Spitalfields mean ages were higher, meaning that if they were aged using Meindl-Lovejoy's standards, systematic underageing (that is, estimated age would be lower than actual age) would occur; while population variation may be the reason for such differences in mean age per stage, the older ages of the Spitalfields individuals compared to Meindl and Lovejoy's (1985) Hamann-Todd Collection sample may also be a reason. The use of as many age indicators as possible has been suggested (Buikstra and Konigsberg, 1985: 318-319; Meindl and Lovejoy, 1985: 65-66; Acsádi and Nemeskéri, 1970: 120) as a solution to this problem of the target sample age distribution mimicking that of the reference sample (Bocquet-Appel and Masset, 1982).

Other tests of Meindl and Lovejoy's (1985) method followed (Hershkovitz et al., 1997; Sabini and Elkowitz, 2006; Sahni et al., 2005), none of which have been particularly supportive of the continued use of cranial suture closure as an age estimation method. Hershkovitz et al. (1997: 398) and Sabini and Elkowitz (2006: 600) suggest that open or partially open sutures may confer an advantage, because these might absorb mechanical stress, as does suture interdigitation, as well as preventing actual separation, thus increasing skull efficiency. The general conclusion is that suture closure is not an appropriate age indicator (Hershkovitz et al., 1997: 398; Sahni et al., 2005: 199, 204).

2.7.3.5 Summary

Overall, the utility of cranial suture closure as an age estimator does not look terribly promising. Despite much work being done over the last 100 years or so, there is still much disagreement. In some populations, endocranial sutures seem to begin closure earlier than ectocranial sutures, but other studies have found the opposite. There are differences in closure times or frequency in degree of closure by sex in some populations but not others. The majority of studies, except for those who are developing new methods (or revising old methods), seem to indicate that there is only a low correlation of suture closure to age, perhaps only in broad terms – that is, generally, there may be high frequencies of individuals with completely open sutures at very young ages, which decreases with age, and low frequencies of individuals with completely closed sutures at the youngest ages, which increases as age increases. However, enough young individuals with closed sutures and old individuals with open sutures places considerable doubt on cranial suture closure as a useful age indicator.

2.7.4 Pelvic Age Indicators: Auricular Surface

The auricular surface is more frequently preserved in archaeological contexts, perhaps providing more opportunity for its use for ageing adult skeletons than the less well-preserved pubic symphysis (see Figure 2.5). While Sashin (1930) had already examined the sacroiliac joints of 257 cadavers and noted that the appearance of the hyaline cartilage covering the auricular surface changed with age, Lovejoy and colleagues (1985b) were the first to investigate and develop the auricular surface of the ilium as an ageing method. Similarities between the descriptions of cartilage and bony change are clear; for instance, the cartilage covering the auricular surface becomes 'coarsely granular' and irregular, with osteophytes around the joint margins (Sashin, 1930:899). Degenerative changes that increase with age were noted in both sexes but began earlier and proceeded more quickly in males. Ankylosis of this joint was also more frequent in males (Sashin, 1930:904; Stewart, 1984: 194). In females, the ligaments become relaxed and increase in mobility during pregnancy (Sashin, 1930:895); perhaps the increased mobility due to pregnancy prevents later ankylosis in females (Stewart, 1984: 195). Even in later life, with the presence of arthrotic changes, female sacroiliac joints have more movement relative to males (Brunner et al., 1991: 1117). In males, ankylosis tends to occur superiorly, across the weight-bearing line; when ankylosis is present in females, it tends to occur anteroinferiorly, suggesting 'that [this] reflects the movements associated with the functional organization of the female bony pelvis' (Stewart, 1984: 195). In older ages, degenerative changes became more marked, increasingly roughened and irregular, with more coarse, fibrous strands encroaching on the joint space, with more frequent ankylosis (Sashin, 1930: 907-908).

2.7.4.1 Lovejoy et al.

While Sashin made no more specific links with age, Lovejoy et al. (1985b) developed the auricular surface as an ageing method, and described morphological change in five year modal ranges (see Figure 2.6 for examples of young and old auricular surfaces). The method was developed on a sample of over 750 individuals comprising the Hamann-Todd Collection (approximately 500 individuals), the Libben collection of archaeological skeletal remains (over 250 individuals), and 14 forensic cases (Lovejoy et al., 1985b: 17). Two smaller samples (each around 100 individuals, not used in the development of the method) from the Todd Collection were used to test the method. For age estimation of a skeletal sample, it was recommended that seriation should be undertaken before final age estimates are given (Lovejoy et al., 1985b: 26), although this is clearly not possible for isolated skeletons. For auricular surfaces that do not clearly belong in a five-year modal category, Lovejoy et al. (1985b: 26-27) suggest classifying according to the features most indicative of age, with auxiliary features used to refine the estimate.

Figure 2.5. Location of auricular surface (indicated by red square). Right side. Image author's own.

The primary features, most important for assigning age, are billowing and striation (in younger individuals), granularity, densification and irregularity beginning at middle to old adult ages, with increased irregularity, breakdown, marginal lipping, macroporosity and significant periauricular activity in older adults (60 + years). No sex differences were found, although for females with well-defined preauricular sulci, which results in marked changes in the apex and inferior margin, these features should be ignored for age estimation (Lovejoy et al., 1985b: 27). Some inaccuracy and bias was noted in tests on the reserved sample (not used in developing the method), outlined in Table 1 below. Lovejoy et al. (1985: 7) define inaccuracy as 'the average absolute error of age estimation for each individual, without reference to over- or under-aging', and bias as 'the mean over or under prediction' of age. Overall inaccuracy was found to range from 7.3 to 7.8 years, and bias from –3.7 to -0.5 years; the authors concluded that their method was a valuable addition to the age estimation roster, and could improve estimates in forensic anthropological applications when used with other available age indicators (Lovejoy et al., 1985b: 27-28).

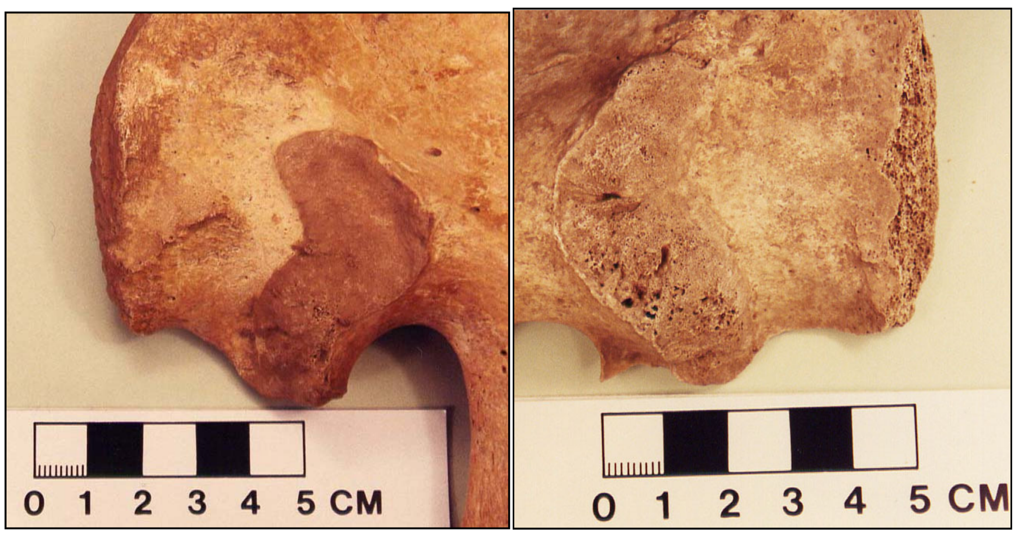


Figure 2.6. Examples of young (left) and older (right) auricular surfaces. Young auricular surface is left side, older auricular surface is right side. Images courtesy of Dr. R. Gowland.

2.7.4.2 Tests of Lovejoy et al.

Independent tests of Lovejoy et al.'s (1985b) method followed. Higher levels of inaccuracy and bias were found compared to that reported by Lovejoy et al. (1985b: 27-28) on samples from the USA (Murray and Murray, 1991: 1167-1168), Italy (Hens et al., 2008: 1043), and Thailand (Schmitt, 2004: 2), while a Canadian sample was found to have similar inaccuracy but higher levels of bias (Saunders et al., 1992: 98,101). Bedford et al. (1993: 290) reported a range of inaccuracy from 9.4

to 11.4 years and a range of bias from -0.1 to 3.4 years, using the Canadian Grant Collection; however, it is worth noting that Lovejoy and Meindl are co-authors of this paper, suggesting that the other observers may have had some instruction from the developers of the method. This possibly lead to bias when comparing these relatively 'good' results with results of other studies whose authors have not benefitted from such instruction of a method that was admittedly 'difficult to master' (Lovejoy et al., 1985b: 15). Murray and Murray (1991: 1167-1168) and Bedford et al. (1993: 291) suggested that the higher frequency of older ages in their respective samples might have been part of the reason for such differences. No significant differences were found by sex or 'race' (Murray and Murray, 1991: 1168; Osborne et al., 2004: 7). Murray and Murray (1991: 1168) suggest that variation was large enough to conclude that the auricular surface should not be used as a single age indicator, but could be used as one of a suite of age indicators.

As illustrated in Table 2.1, overageing (where estimated age is higher than actual age) using Lovejoy et al.'s (1985b) auricular surface method generally occurs until approximately age 39 (although in Bedford et al., it was 59), and underageing begins to occur from approximately age 40 and onwards (Lovejoy et al., 1985b: 27; Murray and Murray, 1991: 1167; Schmitt, 2004: 3; Osborne et al., 2004: 4; Hens et al., 2004: 1042; Bedford et al., 1993: 292; Saunders et al., 1992:104). Inaccuracy also tends to increase with age; often, after 30 to 39, inaccuracy is over 10 years on average (Osborne et al., 2004: 4; Murray and Murray, 1991: 1166; the females in Schmitt, 2004: 3), and by the oldest age group, 60+, inaccuracy can be as high as 31.9 years, as in Schmitt's (2004: 3) male Thai sample.

2.7.4.3 Buckberry and Chamberlain

In 2002, Buckberry and Chamberlain published a revised version of Lovejoy et al.'s (1985b) auricular surface method, tested on an archaeological unknown age sample from Blackgate, Newcastle, for ease of application, and on the known age Spitalfields sample for accuracy. Lovejoy et al.'s (1985b) descriptions of auricular surface change were broken down by morphological feature, so that each feature was scored separately; this was because the changes in each feature appear to be independent of each other (Buckberry and Chamberlain, 2002: 232). Scores are then added together for a component score, which corresponds with a stage and, finally, to an age range with an associated posterior probability (assuming a uniform prior age distribution) (Buckberry and Chamberlain, 2002: 237). Transverse organisation and surface texture are scored from 1 to 5, and microporosity, macroporosity and apical change are scored from 1 to 3 (Buckberry and Chamberlain, 2002: 233-234). The tests on the Spitalfields sample showed that, for this sample, the auricular surface has a higher correlation with age compared to

	Lovejoy et al. (1985b)		Buckberry and Chamberlain (2002)	
	Overage*	Underage**	Overage*	Underage**
Hamann-Todd Collection, USA Lovejoy et al. (1985b)	39	40	--	--
Terry Collection, USA Murray and Murray (1991)	34	40	--	--
Sassari Collection, Italy Hens et al. (2002); Hens and Belcastro (2012)	39	40	59	60
Chiang Mai Collection, Thailand Schmitt (2004)	39	40	--	--
Terry Collection and Bass Donated Collection, USA Osborne et al. (2004)	34	35	--	--
Grant Collection, Canada Bedford et al. (1993)	59	60	--	--
St. Thomas Anglican Church named sample, Canada Saunders et al. (1992)	39	40	--	--
Terry and Huntington Collections, USA Mulhern and Jones (2005)	39	40	49	50
Spitalfields Collection (calculated from data in Buckberry and Chamberlain, 2002), UK Mulhern and Jones (2005)	--	--	59	60

Table 2.1. Ages before which overageing occurs and after which underageing occurs for auricular surface methods
Bias is the mean over- or under-estimation of age: ∑(estimated age – actual age)/*n*
*Overage: the method overages individuals up to X years (given in the table)
**Underage: the method underages individuals from X years and older (given in the table)

the Suchey-Brooks pubic symphysis stages (Buckberry-Chamberlain, 2002: 235-236). The quantitative scoring of each feature was found to be easier to apply than Lovejoy et al.'s (1985b)method (Buckberry and Chamberlain, 2002: 236; Mulhern and Jones, 2005: 65; Hens and Belcastro, 2012: 209.e4), the latter sometimes necessitating forcing a particular auricular surface with varying age features into just one modal category. They also found it possibly more reliable, as age estimates are wider than the 5-year modal stages given by Lovejoy and colleagues; however, this is not related to the availability of casts, because no auricular surface casts are available for either the Buckberry and Chamberlain (2002) or Lovejoy et al. (1985) methods.

Following this publication, studies emerged comparing the accuracy of Lovejoy et al.'s (1985b) method to that of Buckberry and Chamberlain (2002), as well as of the latter alone, on temporally and geographically diverse populations. Mulhern and Jones (2005), using the Terry and Huntington collections, found no sex or "race" differences. Inaccuracy and bias were found to be higher in the younger age groups, and lower from about age 50 and over; however, inaccuracy became higher again after age 72 (Mulhern and Jones, 2005: 64). Conversely, Lovejoy et al.'s (1985b) method was found to be more accurate in the younger and middle adult age groups. Buckberry and Chamberlain's method has been found to provide more accurate older-age estimates, in studies using archaeological Japanese samples, a St. Bride's Church sample, a sample from the Sassari Collection, curated at the University of Bologna, Italy, and a documented modern Spanish sample, curated at the Complutense University of Madrid (Nagaoka and Hirata: 2008; Falys et al., 2006: 510; Hens and Belcastro, 2012: 209.e4; San Millán et al., 2013: 1746, respectively). Furthermore, examination of Buddhist temple records for a slightly later period (from 1812 to 1815) from Toraiwa, an agricultural village in Japan, show an age distribution of deaths that quite closely resembled that produced from the skeletal age estimates using Buckberry and Chamberlain's revised method; this provided further support for the appropriateness of this method for archaeological populations (Nagaoka and Hirata, 2008: 1375). Using older Japanese skeletal data, from the Jomon period (10 000 to 300 BP), Nagaoka et al. (2008: 167) found that Buckberry and Chamberlain's method produced a more realistic age-at-death distribution than the Lovejoy et al. (1985b) method, while Storey (2007), using archaeological Maya skeletons from Late Classic period Copan, Honduras, found the opposite, that the age distribution resulting from Lovejoy et al.'s (1985b) method appeared more realistic (Storey, 2007: 45), despite the overrepresentation of deaths in the 30s and 40s.

Samworth and Gowland, in 2007, used Lovejoy et al.'s (1985b) auricular surface system to develop 'look-up' tables for age estimation, with 90% and 68% prediction intervals. Using the Spitalfields and Coimbra Collections, and Bayesian statistical methods of regressing age against indicator, the tables were derived so that a researcher, after classifying an auricular surface into one of Lovejoy et al.'s phases, need only look on the table under that stage to find an age estimate, including prediction intervals (Samworth and Gowland, 2007: 175-179). The same was done with the Suchey-Brooks pubic symphysis method and the two methods combined. Passalacqua (2010: 483) tested the utility of the look-up tables on samples from the Hamann-Todd Collection, the William M. Bass Collection and the Forensic Data Bank, all American documented collections. While overaging was found at younger ages and underageing at older ages, accuracy ranged from 75% to 99% using the 90% prediction intervals for auricular surface and pubic symphysis together. The conclusion was made that the Samworth and Gowland tables performed adequately on the American samples and were appropriate for use in forensic applications in North America (Passalacqua, 2010: 486).

2.7.5 Pelvic Age Indicators: Pubic Symphysis

2.7.5.1 Todd

As with cranial sutures, Todd was a pioneer in the investigation of the systematic changes of the pubic symphysis with increasing age (see Figure 2.7 for the location of the pubic symphysis). In a series of papers published largely in *the American Journal of Physical Anthropology*, Todd outlined the morphological age-related changes that he observed in the pubic symphyses of white males (1920), black males (1921a), white females (1921b), black females (1921c), and even other mammals (1921d). Other papers on the variation found within pubic symphyses (1921e), the different "strains" of the pubic symphysis (Todd distinguished different "strains" or types of morphological expression, 1923), and whether the morphological age related changes could be seen on roentgenographs (1930) followed.

Figure 2.7. Location of pubic symphysis (indicated by red square). Left side. Image author's own.

The same collection from Western Reserve University (now known as the Hamann-Todd Collection) was used to examine the pubic symphysis as was used in the cranial suture studies, with the same methods of verification of known ages (Todd, 1920: 289-291, 314). Todd identified ten phases of morphological change, with earlier phases (and ages) possessing the ridge and furrow system of billowed bone, changing with delimitation of upper and lower borders at slightly later ages, the formation of the dorsal plateau and ventral bevelling, and then formation of the ventral rampart. Next came ventral bony outgrowths, followed by complete rim formation, and then lipping of the dorsal and ventral margins and, finally, erosion of the face and ventral margin, and osteophytic growths in the final phase (Todd, 1920: 314; see Figure 2.8 for young and old examples). Age ranges are quite narrow, particularly for the earlier phases (see Todd, 1920: 313-314). Todd noted that age is predicted more reliably from 20 to 40 years at death, and that the pubic symphysis can be valuable even at older ages when other pelvic features are also taken into account. No significant differences were found between black and white males (Todd, 1921a: 24). Using a small sample of females (47 white females and 22 black females), Todd (1921b: 27, 39; 1921c: 40, 48) found some sexual differences. Todd suggested that even if using only the pubic symphysis, age could be estimated to within 5 years or less, and, if more age indicators were available, to within 2 or 3 years or less (Todd, 1923: 288). Despite this perhaps overly optimistic statement, Todd (1923: 288) also believed that '…age prediction is at best an approximation'. Even in this early period of investigation into age estimation methods, the improved reliability of age estimates by using multiple age indicators is mentioned: '…although no claim is made that there is any justification for using the pubis alone as an age indicator in case the entire skeleton is available for study' (Todd, 1920: 314).

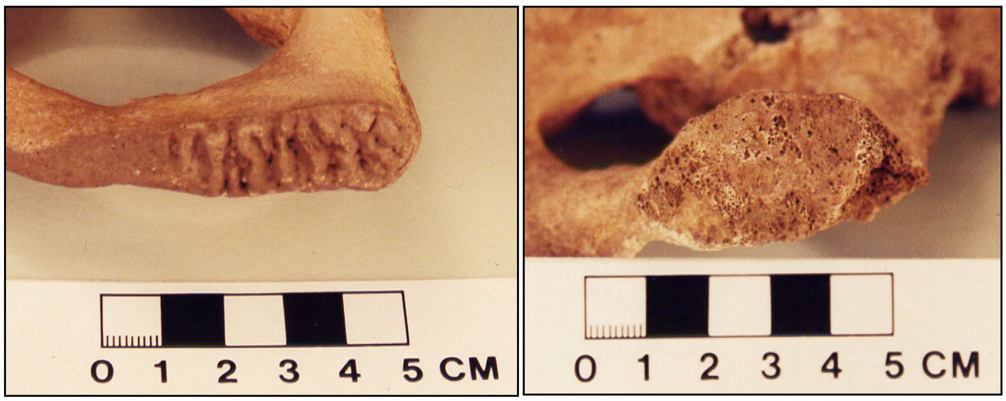


Figure 2.8. Examples of young (left) and older (right) pubic symphyses. Left side. Images courtesy of Dr. R. Gowland.

2.7.5.2 Brooks, Gilbert, McKern and Stewart

Brooks (1955) tested both Todd's pubic symphysis and cranial suture ageing methods, using two archaeological samples of California Indians and a sample from the Hamann-Todd Collection. She found pubic symphysis morphology was more highly correlated with age in males compared to females, and that it was more highly correlated with age than cranial suture closure (Brooks, 1955: 587-588). When tested on the archaeological sample, the resulting age-at-death distribution was found to be unrealistic, with peaks in mortality contrary to those reported ethnologically (Brooks, 1955: 576). Accuracy (known age compared to estimated age) was also found to be poor, lower than that found by Todd (Brooks, 1955: 583). Modified age limits for phases 4 to 9 were presented to increase accuracy, as well as overlapping age limits for successive phases. Interestingly, Brooks (1955: 588) found that Todd's pubic symphysis method resulted in a general trend to overage individuals, with no mention of underageing at older ages. However, upon examination of the scatterplot of known versus estimated ages (Brooks, 1955:

579, Figure 5), it seems that after (known) age 47 or so, individuals were underaged consistently. Unfortunately, not many individuals over this age were analysed – which is not to criticise Brooks, as she specifically set out to use the skeletons that Todd had used in his original study (Brooks, 1955: 570).

Following Brooks' (1955) work, another reworking of Todd's (1920) method was presented, with male and female standards (McKern and Stewart, 1957; Gilbert and McKern, 1973). McKern and Stewart (1957: 72, 73) collapsed Todd's method into a three-component scoring system for age estimation from the pubic symphysis, developed on a sample of American war dead. As such, the age distribution of the sample was skewed towards the younger end, with most ages-at-death between 18 and 24 (McKern and Stewart, 1957: 81). The female standards were developed (Gilbert and McKern, 1973: 34) on a sample ranging in age from 13 to 57 years at death (as no systematic morphological change was found after 55 years), because female symphyses were found to have faster rates of morphological change in the dorsal margin/plateau and slower development of the ventral margin and rampart and symphyseal rim compared to males (Gilbert and McKern, 1973: 36). Some morphological sex differences may be caused by birth trauma – as with the sacroiliac joint, stretching of the ligaments binding the pubic symphyses occurs during pregnancy, and if there is hemorrhage, scars may be left on the dorsal surface of the pubis. Females do seem to display increased variability in pubic symphyseal morphology (or phase compared to age), as demonstrated by the wider age ranges given for each component score, total score and standard deviation in females relative to males (Gilbert and McKern, 1973: 34; McKern and Stewart, 1957: 85). However, that the oldest age ranges are wider for females is hardly surprising given the presence of older aged females; the young war dead sample used by McKern and Stewart (1957: 81) had an oldest age category of 40 to 50 years at death. A later test of the male and female standards on the Terry Collection found that the female standards produced age estimates 'within useful limits in all age ranges', while the male standards, applied to the same female sample, produced surprisingly good results, with acceptable age estimates between 17 and 40 years (Gilbert, 1973: 39). Over 40 years of age, the male standards underaged the female pubic symphyses (Gilbert, 1973: 40).

2.7.5.3 Katz and Suchey

After the McKern-Stewart and Gilbert-McKern standards came into more popular use, tests by other researchers found low accuracy and problems in identifying individuals older than age 59 (Suchey, 1979: 469; Katz and Suchey, 1986: 431). Suchey (1979: 470) also suggested that experience played a role in accuracy of age estimations (Suchey, 1979: 470).

In response to these problems, and to difficulties in differentiating between certain morphological stages (Suchey, 1979: 470), Katz and Suchey (1986: 429, 431) collapsed some of the phases and component scores for each system, developing their new phase categorisation using a sample of 739 males of various ethnicities from autopsies in the County of Los Angeles, California, USA. While the age-at-death category with the highest frequency was 20 to 29, there was good representation up to 79 years at death, with 11 individuals of 80 years and older. Performance was improved when individuals with advanced patterns of morphology were removed, although the authors noted that this was not a feasible technique with regard to forensic applications. Overall, the best results were found using either Todd's original system or Katz and Suchey's modified six-phase version of Todd's system; for easier application, they suggested using the modified six-phase version , for which a table with mean ages, standard deviations and age ranges was provided (Katz and Suchey, 1986: 434).

Following refinements in the description of morphology for the six modified Todd phases, Katz and Suchey (1989: 170-171) renamed the phases to I to VI of the Suchey-Brooks pubic symphysis method, and tested for the effects of 'race' on estimated age, using individuals categorised as 'white', 'black' or 'Mexican' from a Los Angeles, USA, autopsy sample. Overall, significant differences were found in terms of age for pubic symphyses displaying advanced patterns of morphology (phases IV, V, and VI). Specifically, black and Mexican pubic symphyses with advanced patterns tended to be of lower age than white pubic symphyses with the same advanced patterns. However, if only a pubic bone was available, there are no distinguishing morphological features that would allow 'race' to be inferred (Katz and Suchey, 1989: 170). As such, where 'race' cannot be inferred by other skeletal features, Katz and Suchey (1989: 171) note that an estimated age range would have to encompass that of all 'racial' groups. Casts to illustrate features typical for each age phase for males and females were later developed, as well as a set of unisex descriptions (Brooks and Suchey, 1990: 232-233). Brooks and Suchey (1990: 237) believe that the varied ethnicity of the sample used to develop their method is a benefit in terms of application to the diversity of past populations. The importance of using multiple ageing methods whenever possible was also emphasised (Brooks and Suchey, 1990: 237).

2.7.5.4 Tests of Todd, Brooks and Suchey

A number of studies have tested the various pubic symphysis ageing systems (see Table 2.2). These include Klepinger et al. (1992), on the Los Angeles autopsy sample used by Katz and Suchey (1986); Bedford et al. (1993), on the Grant Collection; Saunders et al. (1992), on a St. Thomas Anglican Church sample, from Belleville, Ontario, Canada; Gillett (1991), on an archaeological sample of Californian Indians from a San Francisco Bay site (curated at San Jose State University); Aiello and Molleson (1993), on the Spitalfields Collection; Hoppa (2000), on a Spitalfields sample, and compared to the data of Klepinger et al. (1992); Schmitt (2004), on a documented Thai sample from Chiang Mai (curated at the University of Chiang Mai); Sakaue (2006), on a Japanese sample selected from collections held at the University of Tokyo, Chiba University School of Medicine, Kyushu University and the Kyoto University Museum; Martrille et al. (2007), on the Terry Collection; Djurić et al. (2007), on a Balkan autopsy sample from the University of Belgrade; and Hens et al. (2008), on a sample from the Sassari Collection. While Sakaue (2006: 60, 62-63) found that the Suchey-Brooks method performed acceptably on the Japanese sample, and found fairly low levels of bias (although inaccuracy was higher), others (Klepinger et al., 1992: 766-767; Martrille et al., 2007: 305) found high levels of inaccuracy using the same method, particularly at the older ages (over 60, in most cases) (Saunders et al., 1992: 101, 104; Schmitt, 2004: 2; Djurić et al., 2007: 22; Hens et al., 2008: 1041). Overageing was found to occur at younger ages, and underageing at older ages (Saunders et al., 1992: 101, 104; Aiello and Molleson, 1993: 698; Schmitt, 2004: 2; Martrille et al., 2007: 305; Hens et al., 2008: 1041; Hoppa, 2000: 188). Table 2.2 gives the ages at which over- and underageing occurred. Hoppa (2000: 188) also found differences in rates of pubic symphyseal morphological change between the reference sample and two target samples, but this interpretation suffered somewhat from the fact that Hoppa did not collect all data himself; interobserver differences can thus not be ruled out. Saunders et al. (1992: 101) noted that the age ranges produced by the Suchey-Brooks method are very wide, stating that ‘the estimated ranges are so broad and overlap so much that it is only the very young and very old that are mutually exclusive’. Gillett (1991: 183) found that, compared to Todd’s (1920) method, the Suchey-Brooks method produced a more realistic age distribution at the older ages. Hens et al. (2008: 143), Djurić et al. (2007: 22) and Schmitt (2004: 4) recommend the use of population-specific methods for estimating age.

Klepinger et al. (1992: 769) also tested a pubic symphysis method developed by Meindl et al. (1985a), and found their own results somewhat disappointing. It was postulated that when the developers of a method also test the method, the results may indeed be better than when another researcher, not involved in the method’s development, tests the method (Klepinger et al., 1992: 769). Supporting this conjecture are the slightly better results found by Bedford et al. (1993), when bias and accuracy in Meindl et al.’s (1985a) pubic symphysis method were investigated; included as authors in both papers were Meindl and Lovejoy.

Overall, the Suchey-Brooks method seems to have the most empirical support compared to the other methods listed earlier in terms of accuracy, and is considered the most reliable (Garvin and Passalacqua, 2012: 428). Indeed, the Suchey-Brooks method does seem to be the most widely used method, and was listed as the preferred method for North American forensic anthropologists in osteological practice today (Garvin and Passalacqua, 2012: 428).

	Suchey-Brooks method		Todd original method		McKern-Stewart method (M)		Gilbert-McKern method (F)	
	Over	Under	Over	Under	Over	Under	Over	Under
Brooks (1955:579, Figure 5)			~46	~47				
Saunders et al. (1992)	29	30						
Aiello and Molleson (1993)				~45†	--	~45	--	~45
Bedford et al. (1993)								
Schmitt (2004)	39	40						
Sakaue (2006)	49	50						
Martrille et al. (2007)	40	41						
Hens et al. (2008)	39	40						
Lungmus (2009)	~45 (M) ~52 (F)	~50s (M) ~50s (F)	~46	~47	~29	~30	~53	~54

Table 2.2. Known ages at which bias occurs when using the pubic symphysis for age estimates
Bias is the mean over- or under-estimation of age: ∑(estimated age – actual age)/*n*
Over: the method overages individuals up to X years (given in the table)
Under: the method underages individuals from X years and older (given in the table)
†Aiello and Molleson considered the Todd method with Brooks' (1955) suggestions.

2.7.6 The Fourth Rib as an Age Indicator: İşcan and Loth

The assessment of the relationship between degeneration at the sternal ends of the ribs and age is relatively recent (İşcan et al., 1984a, 1984b); while Loth and İşcan (1989) state that 'the anatomical position, structure, and function of the rib make it a particularly good site to observe the effects of age', these factors are not further discussed. For the development of this ageing method, İşcan et al. (1984a: 1095) sampled 118 white males (of known age, sex and "race") who were autopsied at the Broward County Medical Examiner's Office (Florida, USA), and classified ribs into nine phases (0 to 8). These phases were based on degenerative changes at the costochondral junction, including development of a pit, changes in pit depth and shape, 'configuration' of the walls and rim around the pit, and the quality and texture of the bone (İşcan et al., 1984a: 1094, see Figures 2.9 and 2.10).

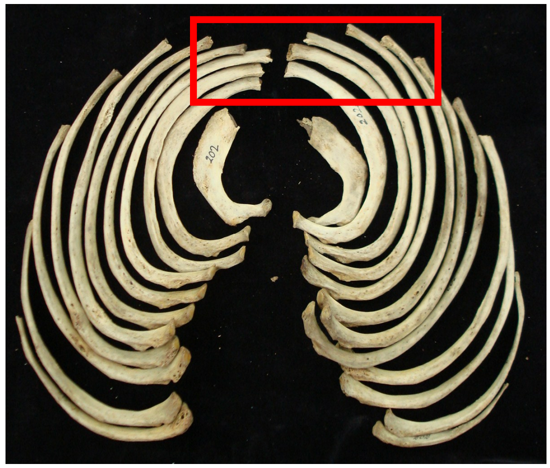


Figure 2.9. Location of sternal ends of the ribs (indicated by red square). Image author's own.

The authors used statistical analysis to confirm that these morphological changes were related to age; they found the quickest and most uniform change occurred across phases 1 to 4, corresponding to ages 17 to 28. Prior to age 17, no morphological change in the sternal end of the rib (other than size) was noticed; as such, this technique is limited to those aged 17 and older at death, and those younger are classified as phase 0 (İşcan et al., 1984a: 1095). After phase 5 (corresponding to age 39), an increase in variation resulted in wider age ranges. The majority (32%) of the sample were in their 20s at death (İşcan et al., 1984a: 1096). The fairly small sample sizes used in the development of this age indicator may have resulted in too-narrow age ranges per phase.

Females were expected to require a separate standard for fourth rib morphological phases to estimate age, as differences in timing of ossification of the cartilage between the sexes was known (leading to the suspicion that sex differences in the timing of morphological change in the bone would also occur), and this was found to be the case (İşcan et al., 1985: 854, 860). Right fourth ribs from 86 white females of known age and sex were sampled, from the same Medical Examiner's Office; the ages-at-death of individuals clustered in the 20s and 40s, with 40% over the age of 50 (İşcan et al., 1985: 854). Females presented evidence of earlier morphological changes than males, starting at age 14 (İşcan et al., 1985: 854, 861-862), as well as having more central bony projections and thinner, less dense bone; however, it is difficult to judge the significance of the early changes in females, as the number of non-adults analysed was not large. It is also not clear exactly how many individuals around these important ages (around 14 to 17) were analysed; for males, the frequency distribution states that ten individuals aged 0 to 16, and six individuals aged 17 to 19 were analysed, and for females, two individuals aged from 0 to 10 years, and six individuals aged between 11 and 19 years were analysed (İşcan et al., 1984a: 1096).

Indeed, the paper later states that only one female between the ages of 10 and 15 was examined (İşcan et al., 1985: 861), so it seems that the age at which morphological change begins must be taken with some caution. Casts illustrating each phase are available for this method, for white males and females only (İşcan and Loth, 1993).

A further exploratory study was undertaken by İşcan et al. (1987) to determine whether there was sufficient morphological variation in black individuals (compared to the white individuals the method was developed on) to warrant separate "racial" standards. Despite a small sample size (63 individuals, with 49 males and 14 females), the authors found evidence of a differing rate and expression of morphological change between black and white individuals, leading to the recommendation of separate standards for these groups (İşcan et al., 1987: 453, 462-464). It is noted that bone remodelling, and hence, ageing, is affected by physical activity, cultural and socioeconomic differences, diet, use of alcohol and/or drugs, and endocrine function, and in this sample, over 50% of black individuals had occupations involving more physical activity than the white individuals, likely causing some of the differences in rate of rib ageing (İşcan et al., 1987: 464).

2.7.6.1 Tests of İşcan and Loth

The Spitalfields collection was later used as an independent sample on which to test this method (Loth, 1995). The ribs of 74 adults were examined, and sex was determined by visual inspection of the ribs (which is not a formal method, but based on the author's experience) (Loth, 1995: 466). The rib seems to have performed well for this site, producing an age distribution fairly close to the actual age distribution for the set of individuals sampled (Loth, 1995: 467), although Molleson and Cox (1993: 174) found that the method performed poorly on Spitalfields individuals. However, although Loth's distribution included individuals placed in the wrong decade of age-at-death, these inaccuracies were well-distributed, resulting in a realistic age distribution. Inaccuracy was lowest in the youngest and oldest age groups, and there was an overall trend towards underageing (Loth, 1995: 467-468). However, though the author states that there is 'consistent underageing' (Loth, 1995: 468), the table presenting the bias for each age group actually shows that for males, the youngest (18-29) and oldest (60-69) groups were underaged, while the other three groups were overaged, and females were overaged at the youngest two age groups and underaged for the remaining three older groups (Loth, 1995: 467). This becomes problematic later, when Loth (1995: 468-469) suggests that the 'consistent underaging' may be due to a slower rate of maturation in archaeological samples (compared to the modern sample the method was developed on); while maturation and ageing rates may indeed have been slower in archaeological populations, the evidence here, of variously under- and overageing, does not seem to support this.

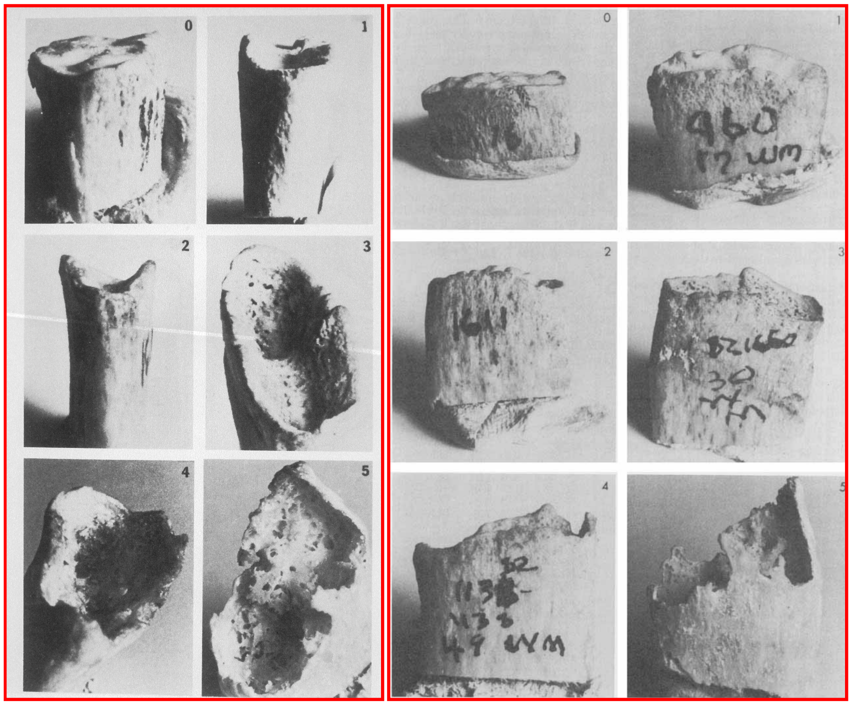


Figure 2.10. Sternal ends of the fourth rib. Left shows pit shape configurations, right shows rim and wall configurations. Pit depth differences can also be seen in the left image. The numbers indicate stages 0 to 5, where 0 represents the youngest and 5 represents the oldest. From Figures 2 and 3, İşcan et al., 1984b: 150-151.

Usefully, Loth (1995: 468) suggests that the third or fifth rib could be substituted for the fourth, when the fourth rib is missing or the actual rib number is uncertain, and other studies have supported this (Dudar, 1993: 796; Yoder et al., 2001: 226; Aktas et al., 2004: 270). It was also suggested that ribs two through nine could potentially be used, with caution, if the fourth rib was not present, or if the rib number could not be identified (Dudar, 1993: 797).

Several independent tests of the fourth rib indicator have been published, on a sample from the archaeological St. Thomas cemetery, Canada (Saunders et al., 1992), on the Hamann-Todd Collection (Russell et al., 1993), on an autopsy and medical school cadaver sample of black South Africans (Oettlé and Steyn, 2000), on a Turkish autopsy sample (Yavuz et al., 1998), on the Terry Collection (Martrille et al., 2007) and on a forensic sample from Arizona, USA (Hartnett, 2010b). Contrary to İşcan et al. (1987: 462-464) and Oettlé and Steyn (2000: 1072), Russell et al. (1993: 57) found no significant variation between white and black individuals; there was, however, a non-significant tendency for the fourth ribs of black individuals to be underaged relative to white individuals. Despite slight differences between the Turkish sample ageing rates and the original American sample, Yavuz et al. (1998: 50, 53) concluded that the fourth rib standards developed on white Americans could be used on Turkish populations, despite geographic, cultural, socioeconomic and genetic differences. Generally, inaccuracy was found to be higher than that reported by İşcan et al. (1984a) (Saunders et al., 1992: 104; Russell et al., 1993: 58; Oettlé and Steyn, 2000: 1073; Hartnett, 2010b: 3-4; Martrille et al., 2007: 304), with a tendency to underage (Russell et al., 1993: 58; Oettlé and Steyn , 2000: 1072-1073). Saunders et al. (1992: 104) and Martrille et al. (2007: 306) found that the method overaged the younger age groups (under age 40) and underaged the older age groups, the pattern typically found with other age estimation methods, as discussed earlier in this chapter. While Saunders et al. (1992: 115) express some concern over the use of the sternal end of the fourth rib due to its tendency to be damaged in archaeological and forensic contexts, Russell et al. (1993: 61) conclude that the fourth rib method can be used as part of a multifactorial age estimate.

2.7.7 Multiple Methods / Multivariate Estimates

Many researchers have advocated the use of multiple skeletal indicators for age estimates in order to improve accuracy and perhaps precision of age estimates, beginning at least as early as Todd, in the 1920s (Todd, 1920: 314; Todd and Lyon, 1924: 380; Acsádi and Nemeskéri, 1970: 120; Meindl and Lovejoy, 1985: 65-66; Brooks and Suchey, 1990: 237; Buikstra and Konigsberg, 1985: 318-319). Indeed, evidence has suggested that the use of multiple ageing methods does increase the accuracy of age estimates (Lovejoy et al., 1985a: 8; Bedford et al., 1993: 296-297; Aiello and Molleson, 1993: 702; Nagar and Hershkovitz, 2004: 153; Baccino et al., 1999: 936; Martrille et al., 2007: 305). However, some evidence is contradictory: Schmitt et al. (2002: 5), in a study using the Dart, Spitalfields, Hamann-Todd and Coimbra collections, a Spanish collection, a Swiss collection and a Thai sample from Chiang Mai, found no improvement by combining age estimates from the pubic symphysis and auricular surface methods when applied to over 900 individuals.

While many have recommended the use of multiple methods, two more formal multivariate methods have also been presented (Acsádi and Nemeskéri, 1970; Lovejoy et al., 1985a). Acsádi and Nemeskéri (1970: 122-135) developed the complex method on a sample of 105 known age Hungarians, involving analysis of age changes in the cancellous bone of the proximal humerus and the proximal femur, pubic symphysis, and cranial sutures. For each method, the mean of each age range, or the upper or lower limit (depending on whether the individual is a young, middle or old adult) is added together, then divided by four for a simple average, giving the overall age estimate (with an increasing range given for increasing means). The complex method has been criticised on a few grounds; namely, the age and sex distribution, socioeconomic status and "race" of their small sample was not provided; their method seems to only differentiate between young and old, and that simple averaging is not an appropriate way of combining multiple age ranges (Brooks and Suchey, 1990: 234, 237; Saunders et al., 1992: 115; Lovejoy et al., 1985a: 3-4).

Lovejoy et al. (1985a: 3-4) developed the summary method of age estimation, using pubic symphyseal and auricular surface metamorphoses, cranial suture closure, dental wear, and age changes in cancellous bone of the proximal femur. The method was introduced and later tested in a paleodemographic study of the Libben site (Ohio, USA), a Late Woodland native American ossuary dating from between AD 800 and 1100 (Lovejoy et al., 1977; Mensforth and Lovejoy, 1985; Lovejoy et al., 1985a), and on an ethnically-diverse sample of 512 individuals from the Hamann-Todd Collection. Seriation was recommended to decrease intraobserver error. After age indicators were assessed, Principal Components Analysis was used to weight the indicators (Lovejoy et al., 1985a: 7-8), producing an age estimate. A 'clinical' age assessment was also included, where other minor age criteria commonly used by osteologists were included to modify age assessment. This included degenerative joint disease, or rejecting one of the age indicators on the basis of a pathological condition interfering with age-related changes in morphology (Lovejoy et al., 1985a: 7). The clinical and summary age methods performed better than any of the age indicators alone, although accuracy was better in the younger age groups for all methods (Lovejoy et al., 1985a: 8-9). Bias was low overall for the clinical and summary methods (2.0 and 2.5 years, respectively), although the maximum bias was -21.5 and -22.3 years, respectively (Lovejoy et al., 1985a: 8, 11). Their selection criteria were later criticised, as individuals with more reliably documented ages-at-death had been selected for; critics claimed this reduced the amount of variability in the sample (Katz and Suchey, 1986: 427; Katz and Suchey, 1989: 167). However, Meindl et al. (1990) produced a statistical analysis of their sample to show that it included an appropriate amount of normal variation. Bedford et al. (1993: 296) later tested the summary age method on a sample from the Grant Collection, and found that this method performed better than the single age indicators. While Saunders et al. (1992: 116) criticised the summary method for individual age estimations on the grounds that it would be effective only if all skeletal age indicators used in the original summary method were available to be studied, Kunos et al. (1999: 322) added their first rib age indicator to the summary method, and found it worked more effectively with the new addition. This suggests that there is some flexibility in the summary method.

Interestingly, while Saunders et al. (1992: 116) did support the use of multiple age indicators, they found that averaging the multiple methods used produced similar results to the weighted Principal Components Analysis suggested by Lovejoy et al. (1985a) although, on statistical grounds, they did not recommend simple averaging. Martrille et al. (2007: 306) suggest determining broad age groups (young, middle or old age), and then choosing age indicators that are more reliable for that particular broad age group, rather than indiscriminately using all available methods.

Evidence has found pelvic age indicators to be the most reliable (Nagar and Hershkovitz, 2004: 153; Saunders et al., 1992: 112; Bedford et al., 1993: 293); as such, more weight is often placed on pelvic age indicators in multivariate methods. For archaeological samples, Lovejoy et al. (1985a: 12) found dental wear to be the most reliable single age indicator; however, this requires calibration by population, as there are differences in tooth wear dependent on diet, which varies by population (Molnar, 1971).

Others have proposed or tested multivariate methods that rely not only on the "traditional" age indicators (like the pubic symphysis, auricular surface and cranial suture closure) but, like Lovejoy et al.'s (1985a) clinical method, take into account minor age criteria, including degenerative joint disease and 'osteophytosis of the femur or pelvis'. Experience and knowledge of normal human variation are important for this type of method. Indeed, İşcan and Loth (1989: 37) noted that osteologists often informally 'average' estimates from various indicators, using their experience and expert judgment, to derive an age estimate. This type of method is more overtly subjective; however, it can be argued that even using traditional phase methods, there is inherent subjectivity, as one must decide into which phase to categorise an individual.

A new subjective age estimation method (Milner, 2007; Boldsen et al., 2002; Algee-Hewitt et al., 2008; Weise et al., 2009) has also been investigated, although full details have not yet been published. The basic premise is that osteologists use minor age criteria (as mentioned above) to estimate ages at death, and that this expertise can be included in a more formal method. Such minor criteria can include the presence of spicules (new bone formation) in the intertrochanteric fossa of the femur, lipping of the fovea capitis, "scooping" (thinning) of the parietal bones, thinning of the maxillae at the canine fossae, degenerative joint disease, "shingle-like" ribs, and angularity of the lateral scapular borders. Some of these minor criteria have been mentioned in past literature. For instance, there are early references to thinning of the parietal bones, leading to depressions antero-posteriorly on each side of the sagittal suture (a scooped appearance, with the sagittal suture remaining almost as a keel) (Ferré, 1876: 423-424). This is listed as a characteristic that may be found in the elderly (with examples given of a 90 year old, a 78 year old and an 88 year old, all female), although the age at which this may begin to occur is not surmised (Humphry, 1890: 1; Ferré, 1876: 423-424). The explicit use of expertise in estimates should not be a disadvantage, as there is evidence that the accuracy of at least some age indicators is better when applied by a more experienced observer (Suchey, 1979: 470; Baccino et al., 1999: 935; Saunders et al., 1992; Klepinger et al., 1992: 769). Statistical methods are to be used with this method, as occurs in the Calibrated Expert Inference, where one indicator is used with a non-parametric regression technique, alongside a maximum likelihood procedure, for an age distribution for the sample and individual age estimates (Weise et al., 2009). In the current study, the minor age criteria of the subjective method were observed, but the full statistical treatments have not been performed, as these are as yet unpublished.

2.8 Conclusion

This chapter has reviewed the importance of understanding the curation contexts of documented collections, and the potential limitations in sampling such collections for research. The historical context and descriptions of the methods tested in this research have also been discussed; the methods chosen for testing here reflect their widespread use in bioarchaeology and forensic anthropology. Human variation, longevity and senescence have been discussed as a broad base from which to begin to better understand the complexity involved in variation in rates of ageing and sexual dimorphism, which is the purpose of this research. The next chapter describes the materials and methods used to fulfill this goal.

Chapter 3: Materials and Methods

3.1 Introduction

This chapter describes the skeletal collections studied and the methods used in this research to examine variability in ageing rates and expression of sexual dimorphism in skeletal remains from different geographic locations and time periods. The composition of the research sample overall and of each collection, along with potential problems due to procedures involved in curating or assembling each collection, is also considered. The skeletal indicators analysed for the purposes of this study are also discussed, as are the comparisons made between the samples from the various collections.

3.2 Materials

Documented collections of human skeletal remains, such as those sampled in this research, are invaluable resources for bioarchaeology and forensic anthropology. "Documented" refers to knowledge of age-at-death and sex of individuals in the collection; sometimes, other information is also available, such as names of individuals and their causes of death. Generally, such collections are fairly recent (from approximately the last hundred years), and also fairly rare: curation of skeletons for research requires amenable local laws or donations of bodies to the relevant institutions, access to skeletons, space and permissions to appropriately curate fairly large numbers of skeletons, and a willing and able curator. Skeletons can be obtained in several ways: the donation of bodies to research, often first dissected by medical students, is one possibility; skeletons may also be curated following excavations of church crypts by archaeologists; or, in some countries where burial is temporary, cemetery staff may excavate burials for reclaimation by relatives, and if skeletons are not reclaimed, curation may occur at that point. It is known that documented collections of human skeletal remains are never without bias (often, non-adults and/or females are absent or few in number, for instance), and are not generally representative of the population from which they are drawn (Usher, 2002; Hunt and Albanese, 2005; Komar and Grivas, 2008; Dayal et al., 2009). Due to their relative rarity, contextual and curatorial variability, each of the collections sampled for this research is discussed.

3.2.1 Samples and Sample Composition

Data were collected from the Grant Collection (early 20th C, Canada), Christ Church, Spitalfields (18th to 19th C, England), Lisbon and Coimbra Collections (early 20th C, Portugal), and Dart and Pretoria Collections (20th C, South Africa) (see Bedford et al., 1993; Molleson and Cox, 1993; Cardoso, 2006; Albanese et al., 2005; Dayal et al., 2009; L'Abbé et al., 2005). These samples were chosen to allow comparison within and between geographic locations and over time; attempts were made to contact curators of other collections, with the hope of expanding the geographical limits of this study, but accessibility was problematic or impossible. The Spitalfields collection is the oldest in terms of date, and the only archaeological population of those sampled.

Ten individuals per decade of life, for each sex were selected randomly for each sample; the youngest adults are 20 years of age-at-death for the purposes of this study, to ensure equality in age ranges for each age category. It should be noted that Buikstra and Ubelaker (1994: 44) state that 20 years of age should be used as the point at which individuals should be skeletally considered adult; in general, long bone epiphyses are fused and third molars are usually erupted. It was important that all necessary skeletal elements, particularly the pelvis and skull, were present for the selection of a skeleton; if the number of individuals per decade could not be fulfilled with all elements for each individual present, individuals missing some elements were selected, but preferably those with elements not specifically being examined (e.g. hand or foot bones). Ribs were the least well represented element overall, and preservation of a complete skull was also variable between and within collections. In some cases, individuals missing either a skull or pelvis (but not both) were included, if sample numbers for a particular age group were very low.

As the individuals with the oldest age-at-death in each collection and the numbers of individuals at the highest ages (80s, 90s, 100+) was variable, sample sizes for each collection accordingly vary; for instance, only the Dart Collection lists several individuals as being over 100 years old at death, so the Dart Collection sample size is larger than that of the other collections. The Grant Collection, however, only consists of 202 individuals, including few people in their 20s and 90s and few females, so the Grant sample size was smaller than that of the other collections. The total sample size for all collections visited was 810 individuals. Table 3.1, below, shows the breakdown of sample numbers by collection.

	Total *n* in collection	*n* sampled	% of collection sampled
Spitalfields Named Sample	387	134	34.6%
Grant Collection	202	83	41.1%
Coimbra Collection	505	140	27.7%
Lisbon Collection	699 (450 available at time of research)	146	32.4%
Dart Collection	2596	163	6.3%
Pretoria Collection	290	148	51.0%
Total		810	

Table 3.1. Total numbers of skeletonised individuals in each collection and number and percentage sampled (*n*: number of individuals)

3.2.2 The Grant Collection

The J.C.B. Grant Collection is curated at the University of Toronto, in Toronto, Canada, by the Anthropology Department. Dr. Grant began curating the human skeletal remains that make up the Grant Collection in 1928, after the Anatomy Act and Revised Ontario Statutes of 1937, 1942, and 1946 allowed unclaimed corpses to be accessioned by medical schools (Bedford et al., 1993: 287-288). Medical students dissected the unclaimed corpses, and remains were later macerated for addition to the skeletal collection. The bodies continued to be brought in (mostly from welfare organisations and hospitals) until the early 1950s (Bedford et al., 1993: 288). Some skeletal elements were later used for anatomy demonstrating purposes; notes on some accompanying documents testify to this purpose.

The Grant Collection consists of just 202 individuals; of these, 175 are male and 18 female. The majority of individuals are of European origin, white, and male, and were unclaimed because they were recent immigrants with no relatives nearby, "transients" or migrant workers (Bedford et al., 1993: 288). Apparently, a decision in 1948 resulted in the disposal of (male) skeletons of unverified age; ages were verified by checking stated ages against hospital records, statements given by the individual before death, or vital statistics records (Bedford et al., 1993: 288), although a few discrepancies remain (discussed below).

Many individuals are missing skeletal elements; for example, many have only a few ribs present and/or are missing the calvarium. Possibly due to the small size of the collection, not many published studies have made use of it, despite the lack of other documented skeletal collections in Canada. Researchers who have used the Grant Collection include Bedford et al. (1993), testing the multifactorial aging method (including using the pubic symphysis, auricular surface, and radiographs of the proximal femur and clavicle), Albanese et al. (2008), testing a metrical sex determination method, and Sharman (2004), who explored sex determination using measurements of the clavicle.

The sample examined in this study was constructed with the help of the author's colleague Hope Kron (to permit a truly "blind" sample), who selected ten individuals per decade of life per sex, wherever possible. For example, the collection does not have ten individuals with ages at death in the 90s, so, for this group, as many as possible were examined. As biographical information was printed on the exterior of the boxes, "Post-It" notes were used to cover this information to ensure that the examination was conducted blind. Individuals were chosen randomly, but with respect to the presence of the appropriate elements. While no individual was complete, the most complete individuals were chosen. For decades of life where it was not possible to choose ten relatively complete individuals (with skull, pelvis, ribs, femur), preference was given to pelves and skulls over the other elements; where only one of the pelvis or skull was present, preference was given to the pelvis over the skull, as more pelvic skeletal indicators (for age and sex determination) were being examined.

The total number of individuals used from the Grant Collection is 83, consisting of 18 females and 65 males. While it was possible to examine ten males with ages-at-death in their 30s, 40s, 50s, 60s and 70s, only three males in their 20s and nine in their 80s could be analysed. Thirteen males with ages-at-death in the 70s were examined, as some in the original sample had missing skeletal elements. There were no suitable males with ages-at-death in the 90s. For females, the maximum number of individuals in an age category was five, for the 60s and 70s; only two females in their 20s, 40s, and 90s, and one each with ages-at-death in their 30s and 80s were appropriate for analysis. There were no females with an age-at-death in their 50s; thus, the age structure for the Grant sample in general and Grant females in particular was not uniform as was possible for the other collections. Table 3.2, below, gives a summary of sampled individuals from the Grant Collection.

Age Group	Females	Males	**Total**
20-29	2	3	5
30-39	1	10	11
40-49	2	9	11
50-59	0	11	11
60-69	5	10	15
70-79	5	13	18
80-89	1	9	10
90-99	2	0	2
100+	0	0	0
Total	18	65	83

Table 3.2. Grant Collection sample by age and sex

Some of the “known” ages are somewhat problematic for this collection. Most individuals were accompanied by two documents, a white document (a ‘skeletal report’, which appears to have been photocopied from an original typed page) and an original brown document, which appears to have been kept as a record from the time of dissection to later. This is usually a dissection checklist, but sometimes there is a list of bones stored in the Collection, including the date of dissection. Some skeletons have notes about some elements being sent to another building for demonstrating purposes. These documents contained the name, sex, and age-at-death of the individuals. A typed index card was also included with each individual, listing the individual reference number, age-at-death, sex, cause of death and any missing elements. A yellow index card was sometimes also present, listing the skeletal elements that had not been degreased (these elements were still available for research, but were greasier than treated bones). The brown document changed in form around 1937. Some have notes on pathological conditions, including a note that a female was ‘obese’. The brown documents usually noted the age with ‘- V’ afterwards, presumably an abbreviation for “verified”, as a few ages have a question mark instead of a “V”, and these tend to have inconsistencies in listed age-at-death between the documents.

3.2.3 Christ Church, Spitalfields Collection

The skeletal collection often referred to as “Spitalfields”, or the Spitalfields Collection, consists of excavated skeletal remains from the crypt of Christ Church, in Spitalfields, east London, England. Christ Church is located west of Brick Lane, east of Bishopsgate and Liverpool Street Station (Cox, 1996: 1), and was one of the churches commissioned by Queen Anne under the 1711 Fifty New Churches Act (Adams and Reeve, 1987: 247). Although fifty were commissioned, only 12 or 13 were actually built; Nicolas Hawksmoor designed six of these, including Christ Church, which was built between 1714 and 1729 (Litten, 2002: 221; Adams and Reeve, 1987: 247). Christ Church’s crypt was used for interments between 1729 and 1859 (Adams and Reeve, 1987: 247); the former was the year of consecration of the church, the latter when burials in the crypt officially ceased, although the last burial to actually take place was in 1852 (Cox, 1996: 16). Excavation took place between 1984 and 1986, after a 1981 decision to clear the vaults for installation of necessities such as a boiler room, kitchen and toilets (Cox, 1996). The remains are currently curated at the Natural History Museum in London, under the direction and management of Dr. Margaret Clegg and Mr. Rob Kruszynski. Spitalfields represents the only archaeological sample used in this research, as archaeological skeletal populations are not typically of known age and sex.

While most of the 968 coffins excavated were single interments, some coffins contained more than one individual (Molleson and Cox, 1993: 17). In some cases, one or two infant skeletal elements were found intermingled with adult skeletons, but it is unknown whether the cause was taphonomic processes (coffins collapsing in on each other or decaying away), or whether these represent double burials (Molleson and Cox, 1993: 10). A range of coffin types were used, including wooden, wood-lined lead, and triple layered wood-lead-wood coffins (Molleson and Cox, 1993: 17). The degree of preservation was also highly variable. While many skeletons were in good condition, some were very fragile and crumbly; as noted by Molleson and Cox (1993: 10), some individuals were represented by ‘...a sediment of crystal debris’, while others still had intact soft tissue (including internal organs, skin and hair) or were naturally mummified. Indeed, the author observed the range of preservation, from the crystal debris to a mummified finger in a bag of ribs (and duly notified the curators). Along with the human remains in varying states of decomposition were some grave goods, burial clothing and coffin textiles, insect remains, fungal blooms, adipocere, and minerals (mostly brushite) (Molleson and Cox, 1993; Cox, 1996).

The total number of individuals excavated from the crypt of Christ Church was 987; of these, 600 are unidentified, but 387 are of known name, sex and age-at-death, due to inscriptions on coffin plates (Cox, 1996: 11). This 'named sample' has been widely used by the bioarchaeology research community. The named sample of the Spitalfields collection is invaluable as a documented sample of human skeletal remains from the 18th to 19th C, and studies have included those on growth and development, paleopathology, sex determination and age estimation (e.g. Samworth and Gowland, 2007; Rogers, 2009; Liversidge and Molleson, 1999; Humphrey and Scheuer, 2006; Roberts, 2007; Ali and MacLaughlin, 1991; Mays, 2000; Mays, 2001; Buckberry and Chamberlain, 2002; Key et al., 1994; Loth, 1995; Megyesi et al., 2006; Owers and Pastor, 2006; Lewis, 2002; Rosas et al., 1999; Heuzé and Braga, 2008; Lewis and Gowland, 2007; Sulzmann et al., 2008; Liversidge, 1994; Waldron, 1997; Waldron and Cox, 1989; Wilson et al., 2008; Cox and Scott, 1992; Scheuer and MacLaughlin-Black, 1994; Cowal and Pastor, 2008; Hoppa, 2000).

Collating biographical histories for individuals belonging to the named sample was undertaken by Cox (1996; also see Molleson and Cox, 1993). Records of baptism, marriage and burial were consulted, as were non-church records, such as trade directories, newspapers, personal papers, coroners' reports, and death certificates (Cox, 1996: 13). Information for some individuals includes details of occupation, address, socioeconomic status, and family size. Many of the families of the Spitalfields parish were of Huguenot origin (see below); for these families particularly, living descendants were able to provide further information (Cox, 1996: 13).

Most of the named sample has French surnames (42%), although very few were not born in England (Cox, 1996: 17). Other surnames are English (33% of the named sample) or unclear, although some seem to be from the Low Countries (in modern terms, the Netherlands, Belgium and Luxembourg). The Huguenots were French Protestants who left France to avoid religious persecution between the late 16th and mid-18th centuries (Cox, 1996: 17). A large number of the Huguenots who eventually settled in Spitalfields hailed from the Bordeaux and Saintonge areas of France and were involved in the French silk industry; these refugees brought their trade with them, making Spitalfields the hub of the silk industry in England (Cox, 1996: 17-18). Interestingly, Litten (2002: 21) notes that the luxurious silk and velvets produced in Spitalfields were used in the funerary trade, and was also called the 'black stuff' industry. This is reflected in the listed occupations of the named sample; of the 237 individuals for whom occupation was identified, 40% of these were within the silk industry, including master weavers, journeyman weavers and silk dyers (Cox, 1996: 58-59). Other occupations are diverse, ranging from high status positions such as an MP and a surgeon, to those in construction or the food and retail industries, to low status occupations such as a bird dealer and a brushmaker (Cox, 1996: 58). Many were of the so-called 'middling sort' and artisans (of lower socio-economic status than professionals, merchants or master craftsmen); interestingly, a shift in socioeconomic status is found in Spitalfields with the turn of the 19th century. The 18th century deaths largely reflect master craftsmen, while the next century's dead were more likely to have been artisans, reflecting a change in the socioeconomic status of the population in the Spitalfields area (Cox, 1996: 68-69). In terms of material wealth, artisans were the only group not to own property (Cox, 1996: 68).

The Christ Church Spitalfields sample was chosen randomly by the researcher prior to visiting the collection, from a spreadsheet provided by Dr. Margaret Clegg of the Natural History Museum with details of the named sample. Individuals were chosen by decade of age-at-death randomly, and a list generated of the chosen individuals' skeleton numbers only, to ensure blind testing. Because crania, postcrania and ribs are curated separately for this collection, crania were analysed first, followed by postcrania and ribs. Two of the chosen skeletons could not be located – for these, alternatives were chosen. The skeletons in poor condition (e.g. very crumbly and brittle) were mostly noted on the spreadsheet, and thus were avoided.

The sample chosen consisted of 134 individuals, including 69 females and 65 males; Table 3.3 provides the breakdown by age category and sex. There were three extra males, one in the 30 to 39 years at death category, one in the 60 to 69 years category, and the other in the 70 to 79 years group; this was because one each of the original chosen sample had missing elements. The named sample does not have a large number of individuals with an age-at-death in their 20s; only nine females and six males in this decade were available to sample. Furthermore, only five males were available for examination with ages-at-death in their 80s, and only one in the 90 to 99 age group; there were no females in the named sample in this last age group.

Age Group	Females	Males	**Total**
20-29	9	6	15
30-39	10	11	21
40-49	10	10	20
50-59	10	10	20
60-69	10	11	21
70-79	10	11	21
80-89	10	5	15
90-99	0	1	1
100+	0	0	0
Total	69	65	134

Table 3.3. Spitalfields Collection sample by age and sex

Buckberry and Chamberlain's (2002) auricular surface component method was developed using the Christ Church, Spitalfields crypt population, so it was expected that this method would be most accurate for ageing the Spitalfields sample.

3.2.4 Coimbra Collection

The Coimbra Identified Skeletal Collection (*Colecção de Esqueletos Identifacados)* at the Museum of Anthropology, University of Coimbra , in Coimbra, Portugal, has been used in past research on age and sex estimation, growth and development, dental caries and tuberculosis (Santos, 2000; Cardoso, 2008a to c; Bruzek, 2002; Coqueugniot and Weaver, 2007; Belcastro et al., 2008; Rissech et al., 2006; Santos and Roberts, 2006; Wasterlain et al., 2009; Correia et al., 2005; Albanese et al., 2005; Albanese, 2003a; Rissech and Malgosa, 2007; Santos and Roberts, 2001; Rougé-Maillart et al., 2009). These skeletons are of 505 individuals who died between 1904 and 1938 and were exhumed from the *Cemitério Municipal da Conchada* in Coimbra (Wasterlain et al., 2009: 66; Santos and Roberts, 2006: 38). Exhumations occur regularly in Portugal after interments of five years in order to move skeletons to ossuaries (Wasterlain et al., 2009: 66); it is necessary, though, for relatives to claim the bodies before deposition in the ossuary and pay ossuary fees – failing this, bodies are incinerated or reburied in communal graves (Wasterlain et al., 2009: 66). Professor E. Tamagnini collected the skeletons curated in Coimbra University between 1915 to 1942, instead of cremation or reburial in communal graves (Coqueugniot and Weaver, 2007: 425; Wasterlain et al., 2009: 66). Coqueugniot and Weaver (2007: 426) note that some individuals were unclaimed at their own direction or those of their family, and not purely for financial reasons. The ages-at-death range from seven to 96 years, with years of birth from 1826 to 1922 (Coqueugniot and Weaver, 2007: 426); while the vast majority were born in Portugal, six were African-born, one Brazilian-born and two were born in Spain (Coqueugniot and Weaver, 2007: 427).

Socioeconomic status has been inferred by the occupations of individuals in this collection (Coqueugniot and Weaver, 2007) or, for individuals who died in hospital, from medical records that listed grades of socioeconomic status, with first and second class patients responsible for paying hospital fees, while the third class and poor patients did not (Santos and Roberts, 2001: 40). Most of the females are listed as housemaids or housewives, although a few were seamstresses; male occupations are more varied, and included artisans, “workers”, and soldiers (Santos and Roberts, 2001: 40; Coqueugniot and Weaver, 2007: 427). The majority of individuals are considered to have been of low socioeconomic status (Coqueugniot and Weaver, 2007: 427; Belcastro et al., 2008: 150; Wasterlain et al., 2009: 66).

Wasterlain and coworkers (2009: 66-67) note that the typical diet for Portuguese people in the first half of the 20th C consisted largely of bread, potatoes, and vegetable soups (green and ‘dry’ vegetables). Bread could be made of rye, wheat or barley, but was typically made of maize; while fish was consumed (*bacalhau*, or salted cod, which is still popular today, or sardines), this was not every day. This simple diet probably reflects that over 50% of the earnings of low socioeconomic status Portuguese during this time was used to purchase food for an adequate diet (Cardoso, 2005: 39). For rural dwellers, the figure was higher. While these statistics are listed for people living in or around Lisbon, the inhabitants of Coimbra would not likely have fared much better.

Information regarding each individual has been compiled into a record book, including name, age-at-death, sex, place and cause of death, marital status, occupation for most, name of the parents, birth place and location in the cemetery of original burial. *Freguesia* is also listed; these are how municipalities in Portugal are divided for administrative purposes (Cardoso, 2006: 174). Cause of death information has been correlated with hospital records, autopsy records and other records where possible; a good correspondence between the collection's information and the other records indicate reliable data (Santos and Roberts, 2001: 40). Tuberculosis and other infectious diseases were the most common cause of death listed for these individuals, followed by circulatory disease and heart disease (Coqueugniot and Weaver, 2007: 427).

The Coimbra Collection sample was chosen using the record book, which is kept with the collection. Individuals were chosen at random for each age category until each was full, although Table 3.4, below, shows that for some age categories (e.g. males aged 20 to 29), it was not possible to sample ten individuals. Data were collected from 140 individuals from this collection, including 73 females and 67 males. There is an extra male in the 60 to 69 year category, as one of the original choices was missing key skeletal elements. While it was possible to sample ten individuals for most age groups, there were only nine males in their 20s, six in their 80s, and one in the 90 to 99 age group; only the 90 to 99 years at death group is underenumerated for females, with only three in this category.

Age Group	Females	Males	**Total**
20-29	10	9	19
30-39	10	10	20
40-49	10	10	20
50-59	10	10	20
60-69	10	11	21
70-79	10	10	20
80-89	10	6	16
90-99	3	1	4
100+	0	0	0
Total	73	67	140

Table 3.4. Coimbra Collection sample by age and sex

The skeletons are generally well-preserved, although there is damage of elements for some individuals. Each skeleton is curated in a wooden box with a small plaque on the front showing the individual reference number. The reference numbers are in blue or pink, corresponding to male and female individuals respectively. This was actually unnoticed by the author until approximately halfway through data collection –an attempt was then made to “not notice” the colour. This is a possible source of bias in the skeletal determination of sex. Accuracy of sex determination for the Coimbra Collection was compared to that for the other collections to test for bias (that is, if accuracy of sex determination is higher for the Coimbra Collection, possibly due to colouration of the reference numbers).

3.2.5 Lisbon Collection

The Lisbon Collection, or Luís Lopes Collection, is one of two collections of identified human skeletal remains curated at the Bocage Museum, in Lisbon, Portugal. The Bocage Museum is the Department of Zoology and Anthropology of the National Museum of Natural History (Cardoso, 2006: 173). The Lisbon Collection is the newer of the two; the older collection is the Ferraz de Macedo Collection (or ‘the old Lisbon collection’), which was largely destroyed by a fire in the National Museum in 1978. This older collection was composed mostly of crania, of which few remain (Cardoso, 2006: 173). The newer collection was begun in the 1980s by Luís Lopes to replace the Ferraz de Macedo Collection, but is composed of complete skeletons. Thus far, research and subsequent publications have mostly been by Portuguese researchers, but Cardoso (2006) published a brief communication with the intent of raising awareness of this series of skeletons. Curation is currently being completed and improved, including cataloguing and storage; increased access will follow. Some mixing of elements was noted by the researcher during data collection and Dr. Cardoso was notified, but fixing this problem was one of the goals of his curation work.

The collection includes 1692 identified individuals and 75 unidentified individuals. While most of these skeletons were collected during the late 1980s until 1991, the collection was reinitiated in 2000, focusing mainly on non-adults and young adults to fill in gaps in the series (Cardoso, 2006: 174). Supplementary details including cause of death, age-at-death, date of death, place of birth and residence and occupation are currently available for 699 individuals; these skeletons are available for study. The skeletons were exhumed from three Lisbon cemeteries – Prazeres, Benfica and Alto de S. João – following the aforementioned practice in Portugal of burying individuals in temporary graves for five years or until skeletonization of the bodies (Cardoso, 2006: 174). This allows for reuse of the grave space, which is at a premium. Skeletons are reburied in communal graves or in *ossários*, or block compartments for storing the remains in urns. The *ossários* require regular fees to be paid by relatives; if payments cease, bones are cremated or placed in a communal grave. Instead of cremation or reburial, if no relatives claim the skeleton after a few years, the Bocage Museum curates the bones (Cardoso, 2006: 174). Cemetery records provide basic data on the individuals, although with information from cemetery records, correlations with civil registrations can be made to obtain additional data, such as parents' occupations, *freguesia* of death and other information. Some individuals from the Lisbon Collection currently have more extensive documentation than others, although the documentation process is ongoing (Cardoso, 2006: 174-175, and see Cardoso, 2008a, for further information on occupation of Lisbon individuals).

Most of the individuals in this collection were born in Lisbon and were of Portuguese nationality, although some were born overseas, including the former Portuguese colonies of Mozambique or Portuguese India (Cardoso, 2006: 175). Ages-at-death range from birth to 98 years, and dates of death are between 1880 and 1975; dates of birth (1805 to 1972) have been calculated from age-at-death – it is possible that some ages have been misreported, because literacy in Portugal was low in the first half of the 20th century (Cardoso, 2005: 31). The literacy rate for Portugal for 1860 was only 12% – compare this to the literacy rate for the UK in the same year, at 69% (Tortella, 1994: 11). By 1950, Portugal's literacy rate was 56%, while the UK's literacy rate was 100% in that year. Socioeconomic status has been inferred from male occupations as low to middle class (Cardoso, 2006: 175). The population was urban, and occupations varied, with 30% classed as service and sales workers, and a further 23% as skilled workers, craftsmen or similar; most females were housewives (85%), although some were teachers, maids or students (Cardoso, 2006: 175). Diet was likely simple, with staples of bread, soup and potatoes with some fish, as described above in the section on the Coimbra Collection.

The most common causes of death for people in the Lisbon Collection were circulatory issues (including cerebrovascular accidents and arteriosclerosis), with infectious disease as the second most common cause of death – among these, tuberculosis is the most frequent (Cardoso, 2005: 49). Deaths from cancer are the next most frequent cause of death (Cardoso, 2006: 175).

Previous studies using the Lisbon Collection have included paleopathological research, sex and age estimation studies, and biocultural work examining trends in stature and growth and development (Cardoso, 2005; Cardoso, 2007a; Cardoso, 2007b; Cardoso, 2008a; Cardoso, 2008b; Cardoso, 2008c; Cardoso, 2009; Cardoso and Garcia, 2009; Matos and Santos, 2006; Heuzé and Cardoso, 2008; Cardoso and Gomes, 2008; Cardoso and Saunders, 2008; Matos, 2009; Vlak et al., 2008; Rissech et al., 2007; Rissech and Malgosa, 2007; Rissech and Malgosa, 2005; Rissech et al., 2008; Rissech et al., 2003; Albanese et al., 2005; Rios et al., 2008; Rogers, 2009; Cardoso, 2008).

Most of the skeletons are in a fairly good condition, although there is variation. Soft tissue preservation is not uncommon, including hair and cartilage. While most skeletal elements are present, smaller hand and foot bones are not always completely present. Cardoso (2005: 58; 2006: 175) notes that the cemetery workers who performed the exhumations were not trained in osteology and might have been likely to overlook small elements. Furthermore, washing the skeletons following exhumation is customary, and may lead to some damage. Storage in individual urns in *ossários* (small niches or compartments for each urn) may result in flaking in some cases (Cardoso, 2005: 59); long bones are placed longitudinally inside the urns, and where moisture or water has seeped in, damage is characteristic on the ends of bones that touched the bottom of the urns (Albanese, 2008, pers. comm.)

For this research, the total sample used from the Lisbon Collection was 146 individuals, consisting of 76 females and 70 males; Table 3.5 provides the numbers of individuals in the sample by age and sex. The sample was chosen at random from an Excel spreadsheet listing sex, age and collection reference number for the individuals, kindly provided by Dr. Cardoso. Only nine females were sampled for the 30 to 39 year group, due to lack of availability, i.e. only nine females with ages-at-death in their thirties were available for study. There were seven females aged 90 to 99 at death sampled, but no males in this group. It was possible to sample ten each in all other groups. Each individual is stored in a drawer in a large cabinet, and the collection is spread over two floors of the Bocage Museum; this is because physical improvements to the building are still being made (necessary due to destruction in the 1978 fire).

Age Group	Females	Males	**Total**
20-29	10	10	20
30-39	9	10	19
40-49	10	10	20
50-59	10	10	20
60-69	10	10	20
70-79	10	10	20
80-89	10	10	20
90-99	7	0	7
100+	0	0	0
Total	76	70	146

Table 3.5. Lisbon Collection sample by age and sex

Cardoso (2005: 34) has highlighted the relative homogeneity of Portuguese people, both ethnoculturally and linguistically. He also emphasises that there are ‘no major religious, linguistic or ethnic minorities’ or subnational divisions (Cardoso, 2005: 34), although as Portugal previously had colonies on every continent, genetic contributions from these territories must surely have occurred. While there is more diversity in urban areas, such as Lisbon, this relative homogeneity was tested in terms of ageing rates and sexual dimorphism, by analysing intra-national variation between the Lisbon and Coimbra Collections.

3.2.6 Dart Collection

The Raymond A. Dart Collection is curated at the Anatomy Department of the University of the Witswatersrand, in Johannesburg, South Africa, and consists of 2605 individuals of known age and sex (Dayal et al., 2009: 11). This collection of human skeletal remains was begun in 1923 by Raymond A. Dart, Head of the Anatomy Department at the University. This was because he was inspired by a six month visit as a Rockefeller Fellow to Washington University in the United States, where Dr. Robert J. Terry was collecting human skeletal remains of known age, sex and "race" (the famous Terry Collection, now curated at the Smithsonian Institution in Washington, DC, USA) (Tobias, 1987: 33). While Dart is perhaps most famous for his role as discoverer of the Taung Child, his achievement in the curation of more than 1000 human skeletons of known age and sex by his retirement in 1958 is laudable. Phillip Tobias, Dart's student and successor, named the collection after his teacher and colleague and continued collecting skeletons (Dayal et al., 2009: 325). Maciej Henneberg continued curation in the 1980s and 90s, but focused on ensuring more equality in representation of various groups (for example, by sex or different tribal groups). While curation is ongoing, the numbers of skeletons added to the collection annually is less than seen in previous decades due to an increase in demand by the medical school for teaching anatomy to medical students (Dayal et al., 2009: 326-327).

Skeletons in the collection began as bodies either unclaimed or donated to the Medical School of the University of the Witswatersrand for dissection by medical students (Dayal et al., 2009: 327). There are also some individuals from archaeological sites, victims of mine accidents and forensic cases for whom less information is available (these are not included in the 2605 individuals) (Dayal et al., 2009: 327,330; Tal and Tau, 1983: 215). Maceration processes followed, then degreasing and drying before incorporating skeletons to the collection. The South African Human Tissues Act (No. 65) of 1983 and previous legislation have allowed for medical schools to procure teaching and research material (Dayal et al., 2009: 327).

The collection was initially stored in the Medical School basement, which, in 1959, was flooded after pipes burst in the street outside (Dayal et al., 2009: 326). While staff attempted to rescue the 'free-floating' bones, some mixing of individuals occurred. Skeletons were laid out to dry on the roof, then placed into boxes, but it is known that some mixing of individuals occurred at this point as each element was not labelled with reference numbers at that time (Dayal et al., 2009: 326). Work was undertaken in the mid-1980s to create an electronic record, resolve some problems of intermingling, and to deaccession some skeletons, either due to damage, or lack of provenience (for archaeological or donated/undocumented material). The author noticed some mixing of elements during data collection; the current curator, Mr. Brendon Billings, is working to overcome these issues, but they still represent a possible source of bias.

Another potential source of bias lies in the "known" ages of the Dart Collection. Many of the skeletons originate from unclaimed bodies that were dissected by medical students; furthermore, many were migrant workers from outside Johannesburg (Dayal et al., 2009: 329).

Hospital staff sometimes had to estimate age for the unclaimed bodies – while sex is significantly easier to report when soft tissues are present, age is more problematic, particularly in a racist (apartheid-era) country where many of the dead hospital patients were black and most of the doctors white. Indeed, Tal and Tau (1983: 217) note that up to 25% of the recorded ages differed by over 10 years compared to estimates 'by attrition'. Dayal et al. (2009: 331) tested for age heaping (see discussion of age heaping later in this chapter) by constructing a frequency distribution of listed known ages, and did find heaping at 5- and 10-year intervals. They suggest that many ages ending in 0 or 5 are estimates, but ages in between these intervals are likely accurate. Dayal et al. (2009: 331) advise some avoidance of ages ending in 0 or 5 in assembling research samples; therefore, in choosing the sample for this research, ages ending in other digits were preferentially chosen. A number of individuals have known ages reported as over 100 years at death; all available skeletons listed as 100 or over were examined, and none looked extremely old (although this will be discussed in more detail in later chapters).

The Dart Collection, like many other collections originating from unclaimed or donated cadavers, consists of more males than females; while there are 1840 males, only 756 skeletons are female, for an approximate ratio of 3 to 1 (Dayal et al., 2009: 330; Tal and Tau, 1983: 217). There are also more black South Africans than white South Africans in the collection; South African Asians and Indians and non-adults generally are also underrepresented. There are some individuals classified as 'Mixed', 'Coloured' or 'Hybrid'; "coloured", in South Africa, is an accepted term for a group of South Africans with mixed European, Asian and African ancestry (Dayal et al., 2009: 328). People considered coloured self-identify in this way, but do not seem to identify particular aspects of their mixed heritage, and would simply consider themselves South African and coloured (Mr. Brendon Billings, 2009, pers. comm.). For some individuals, tribal groups are identified (Zulu, Xhosa, and Sotho, for example, are the most common), while others are identified as S.A.N. ('South African Negro'); skeletons of European heritage are variously called 'White', 'Euro' or 'Caucasian'. 'N/S' ("Not specified") refers to black South Africans of unspecified population group. Dayal et al. (2009: 328) note that these discrepancies were caused by South Africa's changing policies on racial classification. There is much variation between the various groups represented in the Dart Collection, in terms of genetics, culture and language; furthermore, apparently even an individual in South Africa will not always consistently identify themselves as being part of a particular 'tribe' or 'race' (Dayal et al., 2009: 327). This inconstancy is apparently due to a separation of the concepts of ethnicity and biology regarding identity. Further possible confusion arises from the fact that a tribe was established by surname or 'other contextual information' for some individuals (Dayal et al., 2009: 327). However, studies on skeletal morphology have found cranial homogeneity between South African tribal groups (deVilliers, 1968: 118). It is also interesting that while the 'White' group has diverse European ancestry, including from the Netherlands, Portugal, Germany, France and the United Kingdom (Patriquin et al., 2002: 105), studies have suggested that the white South Africans show skeletal differences from the “parent” populations and are distinct (e.g. Loth and Henneberg, 1996; Steyn and İşcan, 1998).

There are also some interesting within group trends in terms of representation of age groups. For instance, among South African Whites (SA Whites), the highest numbers of individuals are from 60 to 69, 70 to 79 and 80 to 89 years at death, while for SA Africans, the majority of skeletons are from 30 to 39, 40 to 49 and 50 to 59 years at death (Dayal et al., 2009: 331). A possible cause for this are differences in socioeconomic status and pathways into the collection; for example, the unclaimed bodies of poor migrant black workers dying (fairly young) in hospital as opposed to wealthier, older, white South Africans bequeathing their bodies to science. While age-at-death, sex, population group, and date of death are available for skeletons in the Dart Collection, cause of death is not readily available. This is unfortunate, as a comparison of the causes of death for the younger black South Africans and the older white South Africans in the collection might be illuminating. Today, major causes of death for black South Africans include AIDS, tuberculosis and interpersonal violence, as well as non-communicable diseases, such as cerebrovascular and heart disease – the number of AIDS deaths tend to decline with age, while cerebrovascular deaths tend to increase (Kahn et al., 1999: 435-436; Hosegood et al., 2004: 667); for South Africa generally, HIV/AIDS, cerebrovascular disease, ischaemic heart disease, lower respiratory infections, violence, tuberculosis, diarrhoeal diseases, road traffic accidents, diabetes mellitus, and chronic obstructive pulmonary disease were listed as the top ten causes of death for all ages for 2002 (WHO, 2006: 3). However, individuals who die of violent or accidental causes are not legally allowed to be dissected by anatomy schools (presumably these individuals must be autopsied at the local morgue; Tobias, 1988: 457), although the author noted instances of healed trauma in some of the analysed skeletons.

As the cadavers dissected by the anatomy school, then skeletonised and curated in the Dart Collection, are largely those of unclaimed paupers, it is inferred that socioeconomic status is low (Tobias, 1988: 457). Here, Tobias was referring to black South Africans in the collection. He further notes that the donated cadavers (bodies bequeathed by either the individuals themselves or relatives) more likely represented individuals of higher socioeconomic status (Tobias, 1988: 457). The suggestion was that those dying in state hospitals were likely from the poorest socioeconomic level and that these skeletons may actually act as more sensitive indicators of stress than a sample from the general population. Patriquin et al. (2002: 105) note that while black South Africans were more likely to have been unclaimed and white South Africans were more likely to have been donated to the Dart Collection (and the Pretoria Collection), the donations were often precipitated by inability to pay for burial, and thus that the white South African donations are also likely to have been low socioeconomic status individuals.

The Dart Collection has been used in a number of age and sex determination studies, stature estimation and paleopathological research, and at least one study on “race” determination; the majority of this work has been undertaken by South African researchers (Van der Merwe et al., 2006; Loth and Henneberg, 1996; Loth and Henneberg, 2001; Franklin and Cardini, 2007; Franklin et al., 2007; Stewart, 1984; Dayal et al., 2008; Steyn and İşcan, 1998; Franklin et al., 2008; Barrier and L’Abbé, 2008; Patriquin et al., 2005; Asala et al., 2004; Patriquin et al., 2002; Asala, 2002; Asala, 2001; Steyn and İşcan, 1999; Steyn and İşcan, 1997; Saunders and DeVito, 1991; Tal and Tau, 1983; Tal and Tau, 1984; deVilliers, 1968; Tobias, 1988; Bidmos and Asala, 2005). This may be due to travel distance to South Africa, and likely also owing to the former political situation in the country.

The sample from the Dart Collection provided the largest total number of skeletons analysed in this study, at 159 individuals, including 77 females and 82 males; Table 3.6, below, presents the sample breakdown by age, sex and ethnicity. It was possible to sample ten individuals per sex per age group for all ages except 90 to 99, with nine individuals each for males and females. There were also two females and three males reported to have died at age 100 or more available for study, and these were also examined. The skeletons are generally in good condition, although some had missing elements. Where extra elements were found, Mr. Billings was notified. Postcrania are stored separately from crania. The sample was randomly chosen prior to arrival using an electronic Excel spreadsheet of individuals kindly provided by Mr. Billings. Individuals listed as S.A.N. (‘South African Negro’), or to a specific indigenous group were chosen preferentially over those listed as ‘White’, ‘Euro’ or ‘Caucasian’. As the majority of the collection consists of black South Africans (and indeed, the majority of the country’s inhabitants are black South Africans), the author decided to sample as many of these individuals as possible. At older ages, there were more white South Africans than black, so white South Africans were sampled when necessary. The list also included a checklist of elements; individuals with missing elements according to the list were avoided, although where postcrania were listed as present, they were not necessarily complete. As mentioned above, ages ending in 0 or 5 are somewhat suspicious in terms of verification, so an effort was made to include more ages-at-death ending in other digits, although those ending in 0 and 5 were not completely excluded.

Age Group	Females			Males			**Total**
	Black	White	Coloured	Black	White	Coloured	
20-29	10	0	0	10	0	0	20
30-39	9	0	1	8	1	1	20
40-49	8	1	1	9	1	0	20
50-59	6	3	1	9	1	0	20
60-69	5	4	1	8	0	2	20
70-79	7	3	0	8	2	0	20
80-89	3	7	0	5	2	3	20
90-99	1	5	0	5	3	1	15
100+	1	0	0	3	0	0	4
Total	50	23	4	65	10	7	159

Table 3.6. Dart Collection sample by age, sex and ethnicity

Numbers of individuals sampled with distinct “tribal” groups listed were too small for comparisons to be made, although it would have been interesting to assess homogeneity regarding ageing rates and sexual dimorphism between tribal groups. Here, the black and white South Africans were treated as one group; while black South Africans were preferentially sampled, white South Africans were sampled when necessary, and are concentrated at the older end of the age range while black South Africans are largely young to middle aged (with some older ages), making possible “ethnic” differences impossible to separate from age-related differences. However, the aforementioned work of Loth and Henneberg (1996) and Steyn and İşcan (1998), suggesting that the white South Africans are morphologically distinct from the parent populations, gives hope to the appropriateness of considering black and white South Africans as one population.

3.2.7 Pretoria Collection

The Pretoria Bone Collection is curated in the Department of Anatomy, at the University of Pretoria’s Medical School, in Pretoria, South Africa. The collection was initiated shortly after the Medical School and Department of Anatomy were established in 1942 (L’Abbé et al., 2005: 197). Reorganisation was undertaken in 2000 to allow easier access for research. Like the Dart Collection, the Pretoria Collection skeletons are either bequeathed or unclaimed bodies (again, provided for under the Human Tissues Act of 1983, for tissue transplants and medical research and training), which are then dissected by medical students before maceration and addition to the collection. There are two groups of skeletons in the Pretoria Collection: those that are used for research and those used for student-teaching (as complete or partial skeletons) (L’Abbé et al., 2005: 198). While information on skeletons in the student-teaching part of the collection may not be complete, age, sex, population group (‘ancestry’ and ‘ethnicity’), and date of death is known for individuals in the research collection, and cause of death is listed for most.

While relatives or spouses have only 24 hours after death to remove a body from a public hospital (from the Tshwane Metropolitan Area, including Mamelodi, Kalafong and sometimes the Pretoria Academic Hospital), individuals originally unclaimed can be claimed at any time afterward from the University. There is a claim rate of approximately 5% per annum from the University (L'Abbé et al., 2005: 198-199). Proof of relationship or authorisation from a magistrate must be presented to claim such individuals. Disease-free individuals under 65 years at death may be used for tissue transplants; otherwise, bodies are imparted to the Department of Anatomy (L'Abbé et al., 2005: 199). Individuals are given an accession number (associated with age, sex, population group, cause of death and other information), and then embalmed. Embalmed bodies are kept for a year or two prior to dissection and subsequent skeletonisation (L'Abbé et al., 2005: 199). Curation is ongoing, with 50 to 100 cadavers accepted every year, adding to the research (mainly) and student-teaching subdivisions of the collection (L'Abbé et al., 2005: 198-199).

Crania and postcrania are stored separately. While most crania are complete, some calvaria or mandibles are missing, the calvaria being sawn off at the time of dissection. While over 6000 cadavers have been accepted by the Department and are awaiting dissection and/or subsequent processing, there are 290 complete skeletons, 704 complete skulls and 541 complete postcrania currently available for research (L'Abbé et al., 2005: 197-198). Dates of birth range from 1906 to 1951 (Steyn and İşcan, 1999: 78). As in other collections, males are more numerous than females and, similar again to the Dart Collection, there are more black South Africans than white South Africans. The white South Africans in the collection tend to be older than the black South Africans. While black South Africans are present in all age groups, there are no skeletons of white South Africans in the 10-19 and 20-29 year groups, and only one and four in the 30-39 and 40-49 year groups respectively. The highest numbers of white South Africans are again in the age groups 60-69, 70-79 and 80-89, while black South Africans are more numerous in the groups 40-49, 50-59 and 60-69 (L'Abbé et al., 2005: 203). Unlike the Dart Collection, for most individuals, population groups are simply classified as 'white', 'black' or 'other'. Some black and 'other' individuals have more specific additional identification – some black individuals have noted tribal groups ('ethnicity'), including Zulu, Shangaan, Tswana, Xhosa, although many are listed as 'black' under both ancestry and ethnicity. Those in the 'other' category of ancestry are listed as coloured under 'ethnicity' (*Kleurling* in Afrikaans), with the exception of one individual listed as Arabic (*Arabier* in Afrikaans), and a few with no specific ethnicity noted. Most white individuals are listed as 'white' under 'ethnicity' as well as 'ancestry', although a few have more specific labels – Portuguese, German or Hungarian. It is not clear whether these individuals were from these other countries or if this is noted as an ethnic background.

L'Abbé et al. (2005: 200-202) provide an interesting discussion on reasons for donation of bodies compared to unclaimed bodies in skeletal collections in South Africa. Evidently, black South Africans tend not to donate their bodies or bodies of relatives due to ancestor reverence. Clearly, this is at odds with the high prevalence of black males in both the Dart and Pretoria Collections. Young black males often migrate to cities from rural areas to find work, and are often unable to contact their families. If these young men die in the city, it is highly problematic for hospital workers to find and contact immediate family of the deceased (L'Abbé et al., 2005: 202). Conversely, black females in South Africa typically do not leave their home areas for work, and thus their bodies are more likely to be claimed upon death.

From the above discussion, it can be inferred that the socioeconomic status of individuals in the Pretoria Collection was likely low. Arguments for the low status of individuals of the Dart Collection can similarly be applied to the Pretoria Collection, due to similarities in sources of skeletons. Population groups are also similar; Patriquin et al. (2002: 105) referred to the ancestry of white South Africans in both collections. The close physical proximity of Johannesburg and Pretoria also support the homogeneity of population groups. The cause of death data discussed in the Dart Collection section apply to individuals in the Pretoria Collection, as the data were for South Africa as a whole.

Further similarities between the Dart Collection and Pretoria Collection may lie in a lack of verification of "known" ages. While no publication has noted any suspicions of problems with the known ages in the Pretoria Collection, conversations with Mr. Gert Lewis raised the possibility of similar circumstances: that of estimation of ages of the dead in hospital by doctors. Mr. Lewis is well-acquainted with the collection as he is in charge of reception of cadavers, acquisition, embalming and storage of cadavers, as well as preparation for maceration. A test for age heaping of the Pretoria Collection skeletal data should suggest whether this is possible; if many ages have been estimated, more ages ending in 0 or 5 would be expected, as evidenced by the Dart Collection data.

This collection has been fairly well-used by South African researchers, particularly in studies of sexual dimorphism, but also in research on stature and "racial" skeletal differences (Oettlé et al., 2009; Oettlé et al., 2005; Pretorius et al., 2006; Van der Merwe et al., 2006; Steyn et al., 2004; Steyn and İşcan, 1998; Barrier and L'Abbé, 2008; Patriquin et al., 2005; Steyn and İşcan, 1999; Steyn and İşcan, 1997; Bidmos and Asala, 2005; Patriquin et al., 2002). Comparison of, or combining, samples from the Dart Collection and Pretoria Collection is fairly common, probably due to the close proximity of the University of the Witwatersrand and the University of Pretoria. Not many researchers outside of South Africa have yet published research using the Pretoria

Collection; L’Abbé et al. (2005: 197) hoped to raise awareness of the research potential of the collection, after its reorganisation in 2000.

The total sample used from the Pretoria Collection was 148 individuals, 74 each of males and females (Table 3.7 gives the sample breakdown by age, sex and ethnicity). There were no individuals listed as 100 years or over. For males, it was possible to sample ten individuals for each category except the 90 to 99 year group, where only nine were sampled. For females, only two individuals were available for analysis in the 90 to 99 year group and only nine in the 20 to 29 year group. There are extra individuals (one in each category) in the 30 to 39, 50 to 59 and 60 to 69 year groups, because skulls were missing for one each of the original chosen sample. Otherwise, there are ten individuals in each age category. Individuals were chosen at random prior to arrival in Pretoria, as Dr. Ericka L’Abbé, curator of the Pretoria Collection, kindly emailed an electronic (Excel) list of skeletons available for research. Crania and postcrania are stored separately, and are also spread across three locations (two rooms on the 4th floor of the building, with crania, and the ‘main bone room’ on the 5th floor, with postcrania and some crania). As such, some flexibility was required in the order of data collection – some postcrania were first examined, then associated crania from the same (main) room, then crania from the other rooms and associated postcrania afterwards.

Age Group	Females			Males			**Total**
	Black	White	Coloured	Black	White	Coloured	
20-29	7	1	1	9	1	0	19
30-39	11	0	0	10	0	0	21
40-49	9	1	0	10	0	0	20
50-59	10	1	0	10	0	0	21
60-69	8	3	0	9	0	1	21
70-79	4	5	1	10	0	0	20
80-89	1	9	0	4	6	0	20
90-99	0	2	0	1	3	0	6
100+	0	0	0	0	0	0	0
Total	50	22	2	63	10	1	148

Table 3.7. Pretoria Collection sample by age, sex and ethnicity

An effort was made at the Pretoria Collection to sample black individuals wherever possible, as for the Dart Collection. Due to the composition of the collection, more white individuals were sampled for the older age groups, because fewer older black individuals were available for study.

3.3 Methods

As discussed in sections 2.6.1 and 2.7.2, the methods of age and sex determination chosen for this research reflect their relative reliability and extensive use in bioarchaeology and forensic anthropology (Garvin and Passalacqua, 2012; Mays and Cox, 2000: 118). Furthermore, no specialist equipment or training was necessary, and the methods chosen did not require destruction of the skeletons.

3.3.1 Sex

For the morphological sex indicators, the procedures described in Buikstra and Ubelaker (1994) were followed; a "final" sex determination combined the skull and pelvis results, but in the case of disagreement, the pelvis determination was considered final.

Skull and pelvic morphological and metrical characteristics for sex determination were also recorded (Albanese, 2003a; Phenice, 1969; Buikstra and Ubelaker, 1994; Walker, 2008). Albanese's (2003a) metrical method of sex determination requires measuring (using his terminology) the maximum length of the femur, maximum diameter of the femoral head, femoral epicondylar breadth, hip bone height, iliac breadth, superior pubic ramus length (SPRL) and acetabular-ischium length (AIL) and subsequent input into logistic regression-modelled equations to give the probability of the individual being male or female. Albanese's (2003a: 7) best-fit model (called modification 1) included maximum diameter of the femoral head, femoral epicondylar breadth, hip bone height, iliac breadth and SPRL, although other modifications were provided for different combinations of measurements (should any not be available for particular individuals). For this study, modification 1 was used whenever possible, but when particular measurements were unavailable (due to missing skeletal elements, or damage to a landmark used for measurement), whichever modification allowed the use of all or most of the available measurements was used instead.

3.3.2 Age

Skeletal age indicators examined focus on the pelvic morphology (the pubic symphysis and auricular surface) (Suchey, 1979; Katz and Suchey, 1986; Buckberry and Chamberlain, 2002; Lovejoy et al., 1985b), the fourth rib's sternal morphology (İşcan et al., 1984a; Loth, 1995), and ectocranial suture closure (Meindl and Lovejoy, 1985). Both Buckberry and Chamberlain's (2002) and Lovejoy et al.'s (1985b) method using the auricular surface were tested. Two final age estimates were produced for each individual: the first used only the "formal" age estimation methods listed above, and was called the "overall" estimate; the second final age estimate used the same "formal" age estimation methods, but also used any "informal" age indicators that were present, such as spicules in the intertrochanteric fossa or lipping (or filling in) of the fovea capitis of the femur, thinning and "scooping" of the parietals, "shingle-like" ribs, thinning of the maxillae, angularity of the lateral scapular borders, degenerative joint disease and osteoarthritis. These informal indicators of age were described as subjective indicators by Dr. George Milner and Dr. Jesper Boldsen at a 2007 summer course on paleodemography, at the University of Odense, Denmark, and were mentioned in one of Boldsen's previous publications (Milner, 2007; Boldsen et al., 2002). The intent was that these subjective indicators could be used alongside transition analysis statistical methods to provide age distributions and individual age estimates. Here, however, these subjective indicators were used to provide additional age information, informing the "subjective" age estimate.

The analysis of the informal indicators as predictors of age (or at least as to whether they provided additional information that might improve age estimates) was not part of the original research plan, so while such traits were recorded when observed from the beginning of data collection, their absence was not systematically recorded until data were collected from the Coimbra Collection (and subsequently-visited collections). At that point, the author realised the benefits of systematic recording of absence as well as presence of traits. Thus, for Grant and Spitalfields, the presence of such indicators have added to the analysis of the indicators' value, but the lack of data on their absence means that it is unclear whether the indicators were simply not present or whether the relevant skeletal element was not preserved for observation.

To produce both the overall and subjective estimates, the estimated age ranges from each of the formal indicators were compared. Essentially, the area of overlap of the age ranges from the formal indicators gave the overall estimate; if one particular age indicator gave extremely different results to all other age indicators, then this indicator would not be taken into account, or would be given less weight (informally, no statistical procedure for weighting variables was used). For the subjective method, the informal age indicators were also considered and used to adjust age estimations accordingly – for example, if an age range from the formal indicators was 50 to 65 years, but there were multiple joints affected by osteoarthritis that suggested an older age, the subjective age estimate would be on the higher end of that range, and perhaps extend it, giving a subjective range of 58 to 75 years. For both overall and subjective estimates, slightly more weight (again informally) was given to the pubic symphysis and auricular surface as these are generally considered more reliable, except where these appeared to give inappropriate age ranges (i.e. if a pathological condition was present). The final age estimates were calculated in this manner because that is how the author was trained to estimate ages from multiple age estimation methods; recent evidence suggests that this (or some slight variation) is common practice (Garvin and Passalacqua, 2012).

3.3.3 Pathological Conditions

Pathological conditions and trauma were recorded in case of any interference with age and sex indicators. For instance, any evidence of trauma to the skull was recorded, particularly that which might interfere with the observation of cranial suture closure or sex indicators (in case scores or sex estimations were later questioned). Figures 3.1 and 3.2, below, show partially healed cranial trauma in skeleton 6185 from the Pretoria Collection, making it difficult to observe sphenofrontal suture closure. Osteoarthritis and joint degeneration were particularly recorded, for the subjective age estimates and also for later reference if any age estimation scores from that joint or measurement (for Albanese's method) were suspect. For example, Figure 3.3, below, shows the left pubic symphysis of skeleton 13 of the Coimbra Collection – the location of the eburnation and porosity certainly was not helpful in assigning a Suchey-Brooks phase. Waldron's (2009: 34) operational definition of osteoarthritis was used: 'Presence of eburnation OR at least *two* of the following: marginal osteophyte; new bone on the joint surface; pitting on the joint surface; alteration in joint contour.'

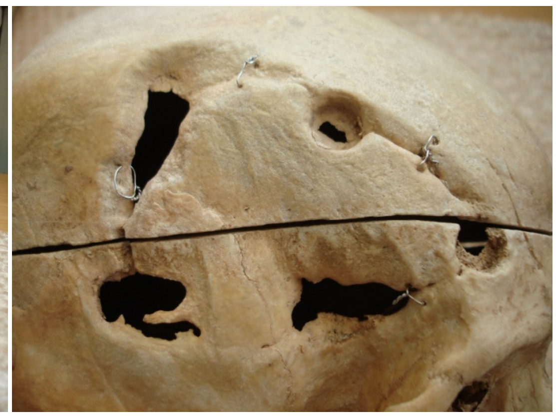


Figure 3.1 (left). Partially healed cranial trauma of skeleton 6185, Pretoria Collection.

Figure 3.2 (right). Closer view of partially healed trauma, skeleton 6185. Images author's own.

Figure 3.3. Left pubic symphysis of skeleton 13, Coimbra Collection, with eburnation and porosity. Image author's own.

3.3.4 Equipment

Equipment used was an osteometric board (travelling) and a digital Vernier caliper (accurate to 0.02 mm), alongside a laptop computer with the database for data recording. The same osteometric board and calipers were used to collect data from all collections for consistency. The Open Office program, Base (an open source version of MS Access), was used to construct the database.

3.3.5 Data Recording

Data collection was blind; known age and sex were recorded and compared to the estimates only after all data were collected. Time spent analysing each individual varied. The most time-consuming part of recording was looking for and describing pathological conditions (as described in section 3.3.3), which generally resulted in a greater amount of time spent examining older individuals because pathological conditions increase with age (in incidence, or in terms of multiple morbidities; e.g. Ortner, 2003: 119, 148, 547, 559; Kalichman et al., 2006). Young individuals tended to have fewer or no pathological conditions. Descriptions of informal age indicators and pathological conditions were recorded in the database if any were present; absence of pathological conditions was not reported (i.e. the 'pathological conditions' field in the database was left blank). Despite time constraints due to travel, all selected individuals were observed fully.

Potential bias may have occurred as the age distribution of the sample was known to the researcher (necessarily, as the researcher designed the project). As such, in estimating age subjectively, an effort was made to not state age ranges strictly within only one decade (for example, estimating a range of 47 to 54 years of age-at-death, instead of 40 to 49). It would be interesting to test a sample completely blind, without any prior knowledge of the age structure of the collection or sample – however, for now, that will remain a possibility for future research.

3.3.6 Other Considerations

Overall, the socioeconomic status of individuals in the collections sampled in this research was low or middle (the individuals of Spitalfields were called 'the middling sort'; see Molleson and Cox, 1993). This can be inferred from the sources of the collections – many skeletons were curated because they were unclaimed by relatives, and may have been new immigrants with no family in the area, transient workers who needed to leave their homes in order to make a living, or fairly old people with no living relatives. In the case of the Portuguese collections, where fees must be periodically paid for remains to be kept in ossuaries after a temporary burial period, financial grounds are often the reason for not claiming a relative's remains (Cardoso, 2005, 2006).

There are, of course, exceptions – individuals may have expressly stated that they wanted to remain unclaimed (Coqueugniot and Weaver, 2007: 426), possibly so relatives did not have to bear the burden of continual payments, and some individuals donate their bodies to science, but these individuals are not necessarily of low socioeconomic status. Thus, within collections, there is likely some variation in the socioeconomic status of individuals. It is also very difficult to compare the lifestyles of the middle classes of 18th and 19th century London with 20th century low status individuals from Europe, Canada and South Africa; variation in many other cultural variables and politico-economic context render such comparison or attempts at calibration inappropriate.

3.4 Data Analysis

For each sex determination method (Albanese, 2003a; Phenice, 1969; Walker, 2008) the allocation accuracy was calculated. For the Phenice (1969) and Walker (2008) methods, where traits are scored individually (ventral arc, ischiopubic ramus ridge, sciatic notch, subpubic concavity for the pelvis, and supraorbital margin, glabella, nuchal crest, mental eminence and mastoid process for the skull), it was also possible to evaluate differences in the score distributions for each trait, to assess variation in sexual dimorphism of each of the traits. The accuracy of the overall sex estimates (using the combination of pelvic and cranial morphology) was calculated. These were then compared between samples from each collection.

For the age estimation methods, accuracy, standard deviation and bias were calculated for each age category (and overall) for age estimates and compared between collections. Mean ages were calculated for each phase or score for each age estimation method in order to compare population ageing rates. Distributions of scores or phases were also examined for each age estimation method. For the Buckberry-Chamberlain (2002) auricular surface method, involving scoring a number of traits separately (transverse organisation, surface texture, microporosity, macroporosity and apical change), then adding these for a component score before a final phase and associated estimated age range are given, it was possible to analyse population differences for each trait. For each trait, mean ages for each score were compared, as was the distribution of scores for each collection. Analysis was done between the collections as well as within collections by sex, to assess whether sex-related differences in ageing were present. The observation of mean age for each phase, alongside associated age ranges for each phase, also allowed the rough assessment of each method for age-related change – that is, if mean age does not rise with phase, or age ranges for the phases are very wide, the relationship of age to the scored morphological skeletal changes, and thus, the utility of the method, is questionable.

3.4.1 Error Testing

An intraobserver error test was undertaken on the Grant Collection; a second round of data were collected from the same individuals to analyse consistency in assessing age-at-death and sex from skeletal indicators. This collection was the first collection visited; a second visit, to undertake data collection for intraobserver error testing, occurred after all other data had been collected. This ensured that if any intraobserver error occurred, it would be the worst-case scenario, as much time had elapsed between visits (about a year), and approximately 700 other skeletons had since been observed. A sample of 20 individuals was taken (24% of the original sample); generally, intraobserver error samples are around 20% of the original sample size. It was not possible to re-sample the other collections for intraobserver error testing due to constraints of time and funding.

Interobserver error testing was undertaken after the kind provision of data by Dr. Rebecca Gowland on the Coimbra and Spitalfields collections. Dr. Gowland had collected auricular surface phase information using the Meindl-Lovejoy method and Buckberry and Chamberlain method, and Brooks and Suchey pubic symphysis phase data. An earlier version of Buckberry and Chamberlain's method was used by Dr. Gowland, where scores began at 0; for interobserver error testing, the scores were adjusted to match the published version (beginning at 1). As it was not certain whether the phase information (from the addition of the individual scores) had changed before publication, only the raw scores for each Buckberry-Chamberlain auricular surface trait were considered, rather than the final phase given. Meindl-Lovejoy and Brooks and Suchey phases do not involve scoring of components, so the phases themselves were compared. From the Coimbra Collection, 68 skeletons were compared for interobserver error testing, while 107 skeletons from Spitalfields were assessed. Comparisons were made of the number of phases or scores that were different between observers, as well as the mean phase or score difference.

Age heaping (as discussed in the Documented Collections section of Chapter 2) was tested for by constructing distribution frequencies of ages-at-death for each collection. As the youngest age-at-death used for skeletons in this research was 20 years old, this was the youngest included age in the frequency distributions. The oldest included age was 114 years, an individual from the Dart Collection, as this was the oldest reported age out of all of the collections, suspicions of the possibility of age exaggeration notwithstanding. Distribution frequencies were constructed with age information for the whole collection for the Dart, Pretoria and Spitalfields collections, but only a partial list was available for Lisbon (the list of skeletons available for research at the time of data collection). For the Grant and Coimbra collections, for which electronic lists were not available, only the samples used here were tested for age heaping. There is no reason to suspect that the samples used here are not representative of the whole collection, as, within the constraints of the age categories, individuals were selected at random.

3.4.2 Statistical Tests

As the main aim of this project was to test for differences in rates of ageing and sexual dimorphism between and within populations, a number of statistical tests were employed for this purpose, using the SPSS (version 19) software package. One-way analysis of variance tests (ANOVAs) were used to analyse differences in mean age per phase or score for each age estimation method. ANOVAs allow testing of means for more than two samples at once. T-tests were also used to examine differences between means, but these are used for only two samples at once. Kolmogorov-Smirnov tests were used to analyse differences in phase or score distribution for each age estimation method between each combination of collections sampled.

ANOVAs are preferable for more than two samples over the use of multiple t-tests, as the results of multiple t-tests increase the possibility of a type I error occurring. For these tests, besides an approximately normal distribution, measurements are assumed to be at least interval level. Other assumptions include random sampling, independent errors, and homogeneity of variance (Thomas, 1986: 255). This assumption of homogeneity of variance, also termed homoscedasticity, ensures that any observed differences between the two samples can be attributed to differences in central tendency, as opposed to different distribution shapes. The null hypothesis is the same again; that the two samples are from the same population.

The two sample Kolomogorov-Smirnov (K-S) test is nonparametric, and essentially looks for abnormalities in the distributions of two categories of independent, ordinal data (Thomas, 1986: 322). There is a one-sample K-S test that is used to check for normality, but this was not used here. As the test is nonparametric, normally distributed data are not a requirement. The samples are sorted into cumulative proportions by variate; in this case, the score or phase (of age or sex determination method) is the variate. This test looks for differences in cumulative distribution, producing the D-statistic, which can then be used to calculate a p-value. The null hypothesis states that the cumulative proportions of the two samples will be ‘essentially similar’ (Thomas, 1986: 322), or are from the same population (Blalock, 1972: 262). Based on this null hypothesis, if p is less than 0.05 at that significance level, the null hypothesis is rejected, and the two groups are different in shape of distribution or location. That is, the two samples may not be of the same population, i.e. there is a significant deviation present somewhere in the cumulative distribution. Here, it is the deviation being tested for; it is known that the different collections represent segments of different populations. The K-S test does not indicate the location of any significant differences that may be present, just whether or not there are any. The K-S test also operates under four assumptions (Thomas, 1986: 324):

- Sampling is random
- Samples are independent
- Measurements are at least ordinal scale
- The underlying distribution of the variables is continuous

Here, the assumption that the data are ordinal is argued; the skeletal data, while scored with either whole or half numbers, are assumed to have a "rank" – a score of 3 represents a more extreme or advanced (age-related) morphology than a score of 1, for example. Certainly, there was a continuous distribution of such skeletal morphology in the samples; sometimes, a particular morphology was fitted into the scoring system of a particular ageing method despite exhibiting characteristics of more than one score, due to the range of morphological variation that is possible.

The Mann-Whitney test, another nonparametric test, also tests whether two samples are from the same population; however, this test checks for differences between the central tendencies of the samples. Again, this test assumes that the two samples are random and independent and that the level of measurement is at least ordinal (Blalock, 1972: 255). It is assumed that the underlying dimension of the data is continuous (Blalock, 1972: 250). The null hypothesis is again that the two samples are of the same form, or are from the same population. The statistic U is calculated by ranking the "scores" of both samples, counting the scores of the first sample with larger ranks, the same for the second sample, and adding these together (Blalock, 1972: 255). If U is very small or large, the null hypothesis can be rejected.

Both the K-S test and Mann-Whitney test were repeated fifteen times for each distribution of score or phase analysed, to compare each collection to each other collection as a pair. Because multiple comparisons increase the risk of a Type I error, the Holm-Bonferroni correction was applied. In this correction, a "new", more conservative p-value is obtained for each ranked p-value by dividing the chosen level of significance (here, 0.05) by *n* minus the ranked number of the pair in terms of degree of significance plus one, where *n* is the number of paired comparisons.

3.5 Summary

This chapter has outlined the materials analysed, the methods of age and sex estimation, the method of recording pathological observations, along with basic methods used for comparing ageing rates and sexual dimorphism. Each collection from which the research sample was drawn has been discussed, including the composition of each collection, source material and biases within each collection. While each collection has its own particular biases, and no collection is representative of the entire population from which it is drawn, by taking into account these biases, comparisons will be possible in order to examine variability in ageing rates and the expression of sexual dimorphism, and thus the shortcomings of using standard osteological methods to estimate age and sex of human skeletal remains in archaeological and forensic contexts.

Chapter 4: Results

4.1 Introduction

This chapter describes the results of age and sex estimation for each skeletal collection, and the data deriving from the statistical tests used to compare the data collected. The data are variously divided by age, sex and collection to compare and contrast the results, and to look for any significant age-, sex- or population-related differences. The relevant aspects of the analysis and comparison of the score distributions for the morphological traits used to determine sex are presented first, followed by sex determination results. In those instances where results were not statistically significant or show no clearly-defined trends, detailed descriptions are provided in Appendices 2 to 9. Sex determination by metrical analysis is described next. The age estimation results follow, which include analysis of the distribution of age phases by collection, as well as mean ages per phase or score for each age estimation method. Results are further subdivided by the age groups used for sample selection, to analyse whether age-related differences are confined to any particular age groups, or if any trends appear in particular collections. The overall and subjective age estimates are next compared. The last sections discuss the results of inter- and intra-observer error testing.

4.2 Sex Determination Using Specific Features of the Skull

The Walker and Phenice sex determination methods involve individual scoring of the morphology of the glabella, supraorbital margin, mastoid process, nuchal crest and mental eminence of the skull, and the sciatic notch, ischiopubic ramus ridge, subpubic concavity and ventral arc of the pelvis, respectively. These individual skeletal features were each tested for variation in score distribution using the two-sample K-S test. As suggested by the name, only two samples can be tested at once, so each collection was systematically tested against each of the other collections. The Mann-Whitney U (MWU) test was also used to test for differences in the median values of the distributions. Collections were tested with the sexes pooled together, females only and males only. The numbers and proportions of correctly-sexed females and males for each collection were also examined by known age group, to see if age plays a role in the ability to determine sex, as suggested by Walker (1995, 2005). Tables A2.16 to A2.33, showing numbers and percentages correct for each age group (by decade), can be found in Appendix 2. The results for each skeletal feature are presented and discussed in turn, with tables including p-values to three decimal places. Significant results are in larger boldface type.

4.2.1 Glabella

Statistical tests were performed for the sexes pooled first, and then males and females separately. Statistically significant differences were found both in score distribution and median for the glabella between Grant and all other collections (see Table A2.1, Appendix 2 for details). The difference in median between Coimbra and Pretoria approached significance. None of the other collections displayed significant differences compared to each other in score distribution.

The sexes were separated, and the K-S and MWU tests were run again. No significant differences were found between any of the collections for females considered alone. A more detailed description of the female score distributions, a table with the K-S p-values, and a table with the distribution of scores (number of individuals and percentages) for each collection can be found in Tables A2.2 and A2.3, Appendix 2.

For female glabella scores, no clear age-related trends were seen; higher scores do not cluster at the oldest age groups. Furthermore, no age-related trends were seen in terms of percentages of correctly-sexed females. The variation in the percentage of correctly-sexed females by age group seems to reflect stochastic, individual-expression related variation instead. Overall, the Grant Collection had the lowest percentage of correctly-sexed females (63.6%), while the other values were fairly close to each other – the highest was for Lisbon (85.3%), but the others followed closely, with 83.3% for Pretoria, 83.1% for both Dart and Spitalfields, and 81.5% for Coimbra.

The males had a number of collection combinations that were significantly different. Significant differences in distribution and central tendency were found between the South African collections (Dart and Pretoria) and every other collection except between Dart and Lisbon, between Pretoria and Spitalfields, and each other; results can be found in Table 4.1.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	**.003**	**.000**	.954	.450	.817	.394	**.001**	**.000**	1.000	.811
Dart			**.000**	**.000**	.021	**.001**	.916	.829	**.004**	**.000**
Grant					.405	.075	**.000**	**.000**	.627	.711
Lisbon							**.003**	**.002**	.331	.371
Pretoria									.017	**.001**

Table 4.1. Glabella, males only: two-sample K-S and MWU results

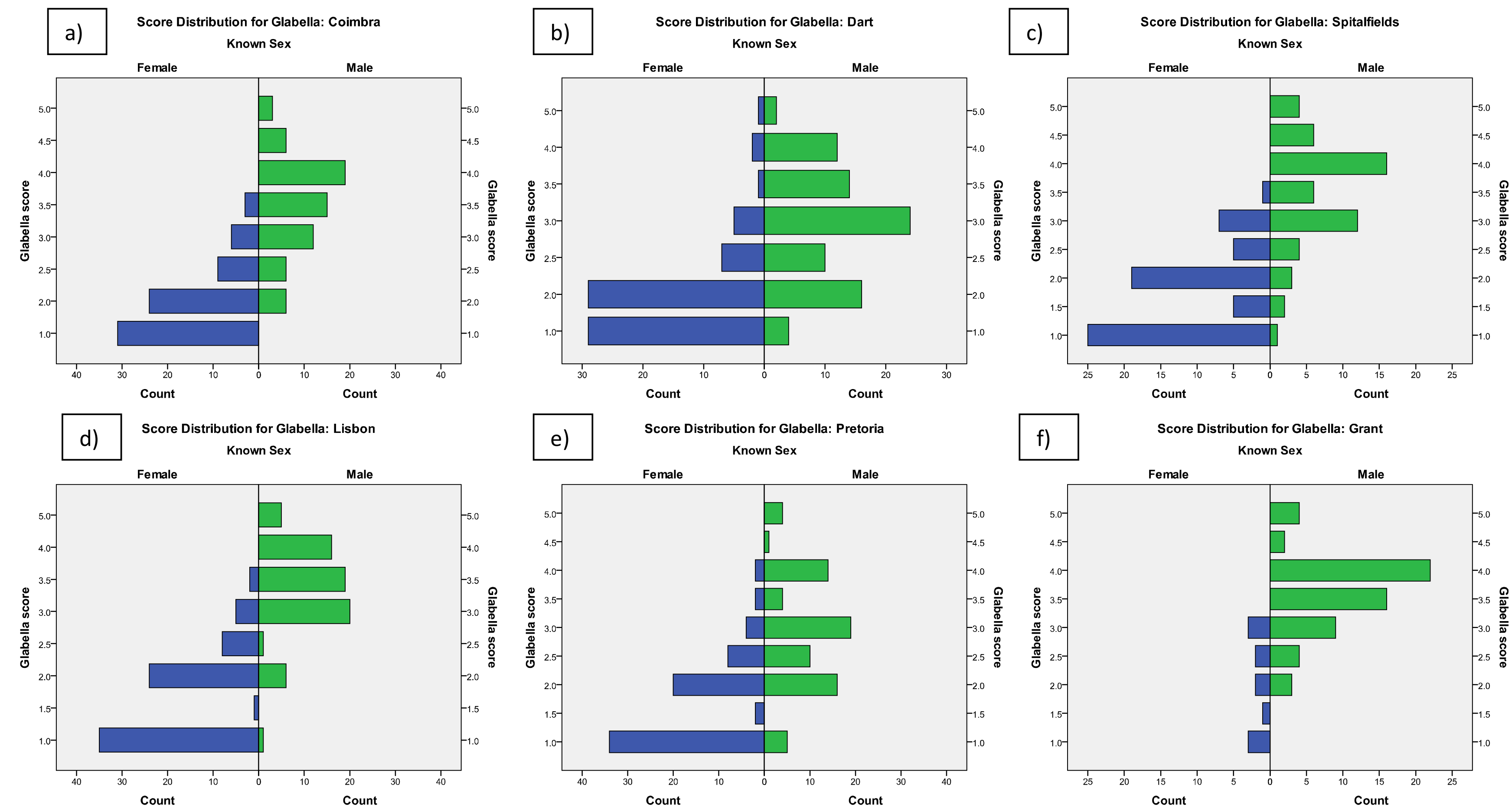
As with the females, the males showed no clear age-related trends in scores; for all except Dart and Pretoria, numbers of males with scores of 1 and 2 tended to be low. However, for Spitalfields, there were slightly higher numbers of individuals clustered in the youngest age groups with lower (morphologically female) scores. For Spitalfields, one individual in the 20 to 29 group, and two from the 30 to 39 group had scores of 2; other scores of 2 included one individual each from the 40 to 49 group, the 60 to 69 group, and 0.5 each from the 70 to 79 and 80 to 89 group. Two scores of 1 were present, from older age groups (distributed from 60 to 89). For Coimbra, the youngest age groups (20 to 29 and 30 to 39) contained about 35% of the morphological female scores (here, only scores of 2); for Lisbon, nearly half of the morphological female scores (1 and 2) were from the youngest age groups. Perhaps a slight bias is evident, but again, no clear trend is visible.

The percentages of correctly-sexed males also did not drop with age – differences again seem to reflect stochastic, individual variation rather than any age-related trend. In terms of overall percentage of correctly-sexed males, it was lowest for Dart and Pretoria (27.4% and 28.8%, respectively), and highest for Grant (60.8%), followed by Spitalfields (55.6%), Coimbra (53.0%) and Lisbon (44.9%). See Table A2.17 in Appendix 2 for full details.

The location of the significant differences from the K-S and MWU tests lie primarily in males from the Dart and Pretoria Collection with scores of 2 – 25.6% and 28.8%, respectively, compared to a mean of 10.6% for the other collections. Grant, Spitalfields, Coimbra and Lisbon also had higher percentages of males with scores of 4 than either Dart or Pretoria. While the percentages of males with scores of 1 were slightly higher for Dart and Pretoria compared to the other collections, the numbers were quite low for all collections. Similarly, scores of 5 were fairly low for all collections, but were slightly lower for Dart and Pretoria. For Grant, Spitalfields and Coimbra, the highest proportions of males had scores of 4, while for Lisbon, Dart and Pretoria, the highest proportions of males had scores of 3. See Table 4.2 and Figures 4.1a to 4.1f for more detail.

Score	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	0	0	2	3.7	0	0	1	1.5	4	4.9	5	6.8
2	5	8.3	6	11.1	9	13.4	6.5	9.6	21	25.6	21	28.8
3	18.5	30.8	16	29.6	22.5	33.6	30	44.1	34.5	42.1	26	35.6
4	31.5	52.5	23	42.6	29.5	44.0	25.5	37.5	20.5	25.0	16.5	22.6
5	5	8.3	7	13.0	6	9.0	5	7.4	2	2.4	4.5	6.2
Total	60	100.0	54	100.0	67	100.0	68	100.0	82	100.0	73	100.0

Table 4.2. Number of males with each glabella score, by collection



Figures 4.1a to 4.1f. Glabella score distribution bar charts for each collection, separated by sex

4.2.2 Supraorbital Margin

The tests for differences in score distribution and median for the supraorbital margin for the sexes pooled found statistically significant differences in both. The MWU tests showed significant differences between the Grant Collection and all other collections except Coimbra, and between Coimbra and Dart, and Coimbra and Lisbon. The K-S tests showed significant differences in score distribution between the Grant Collection had and all other collections except Coimbra and Spitalfields. Table 4.3, below, gives the full results of these tests.

For females alone, significant differences in median were found between Coimbra and Dart and Coimbra and Lisbon. Results can be found in Table 4.4, below.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.034	**.001**	.064	.008	.015	**.000**	.123	.021	.983	.314
Dart			**.000**	**.000**	.998	.613	.638	.210	.145	.026
Grant					**.000**	**.000**	**.000**	**.000**	.032	**.001**
Lisbon							.272	.100	.244	.012
Pretoria									.371	.257

Table 4.3. Supraorbital margin, sexes pooled: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.043	**.003**	1.000	.496	.043	**.001**	.282	.058	.957	.423
Dart			.292	.021	1.000	.662	.620	.144	.213	.046
Grant					.292	.017	.869	.082	.875	.264
Lisbon							.620	.068	.213	.022
Pretoria									.648	.401

Table 4.4. Supraorbital margin, females only: two-sample K-S and MWU results

When female score distributions were examined by age, no clear age-related trends were observed. Scores tended to be equally distributed throughout the age distribution; higher scores did not cluster with older ages. When the percentages of correctly-identified females were considered by age group, again, no clear age-related trends were seen. The overall percentages of correctly-sexed females ranged from 27.3% for Grant to 58.7% for Lisbon, followed closely by the Dart Collection, at 57.4%; 37.0% of Coimbra females were correctly sexed using the supraorbital margin alone, while for Spitalfields, the same value was 43.8% and for Pretoria, 46.5%. Full details are in Table A2.18 in Appendix 2.

When the female score distribution was examined (Table 4.5), the differences between Coimbra compared to Dart and Lisbon were seen. Coimbra females had more scores of 5, as well as fewer scores of 1. Coimbra females also had higher percentages of scores of 3 and 4, and lower percentages of scores of 2 compared to Lisbon and Dart. Scores of 3 represented the highest proportion of scores for Grant, Spitalfields, Coimbra and Pretoria; for Lisbon and Dart, scores of 2 represented the highest proportion. The central tendency values for Grant were higher than others due to a higher proportion of scores of 3 and a lower proportion of scores of 2 (but female sample sizes were quite small); even Coimbra, with its significantly different distribution from Lisbon and Dart, had a slightly higher proportion of scores of 2 and lower proportion of scores of 3. Scores of 5 were uncommon for all collections. Scores of 1 were also not particularly common, ranging from 0% (Grant) to 13.3% for Lisbon. Proportions of scores of 4 were slightly higher in general, but were still low, ranging from 0.7% for Pretoria to 9.1% for Grant (although this represents only one individual). See Figures 4.2a to 4.2f for bar charts of the score distributions.

Score	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	0	0	3.5	5.5	4	5.5	10	13.3	6	8.1	2	2.8
2	3	27.3	24.5	38.3	23	31.5	34	45.3	36.5	49.3	31.5	43.8
3	6	54.5	28	43.8	36.5	50.0	28	37.3	30	40.5	38	52.8
4	1	9.1	5.5	8.6	6	8.2	2	2.7	1.5	2.0	0.5	0.7
5	1	9.1	2.5	3.9	3.5	4.8	1	1.3	0	0	0	0.0
Total	11	100.0	64	100.0	73	100.0	75	100.0	74	100.0	72	100.0

Table 4.5. Number of females with each supraorbital margin score, by collection

For males only, significant differences were found in median between Dart and Grant, and Grant and Lisbon. Further significant differences in score distribution were found between Dart and Grant; all results are presented in Table 4.6, below.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.071	.006	.895	.593	.372	.012	.620	.100	.899	.466
Dart			**.002**	**.001**	1.000	.922	.819	.265	.108	.051
Grant					.027	**.002**	.062	.024	.681	.180
Lisbon							.907	.366	.454	.087
Pretoria									.704	.377

Table 4.6. Supraorbital margin, males only: two-sample K-S and MWU results

When the male score distribution data were examined by age, there were again no clear age-related trends. Coimbra was the only collection to show a very slight clustering of lower scores at younger ages – all together, three of nine scores of 2 (morphologically female) were in the 20 to 29 and 30 to 39 age groups, while the remainder were spread over the rest of the age distribution. However, this still does not necessarily support the argument that younger males display morphologically female skeletal features. In terms of percentages of correctly-sexed males, there was again no clear age-related pattern. Younger or older age groups did not have consistently low or high percentages of correctly-sexed males. Overall, using the supraorbital margin alone, the percentage of correctly-sexed males ranged from 20.1% for the Dart Collection to a high of 50.0% for the Grant Collections. Spitalfields had a total of 38.2%, Coimbra, 40.3%, Lisbon, 25.7% and Pretoria, 29.5%. For full details, see Table A2.19 in Appendix 2.

When the score distributions were examined, the location of the differences between the Grant Collection and the other collections (but particularly Lisbon and Dart, which were statistically significant) were seen. The Grant Collection had lower proportions of scores of 2 and 3, and a higher proportion of scores of 4 compared to the other collections. Indeed, the scores of 4 represented the most males for the Grant Collection. For all other collections, scores of 3 were the most common, ranging from 45.5% (Spitalfields) to 59.8% (Dart). The other collections had higher proportions of scores of 2 and lower proportions of scores of 4 compared to Grant. Scores of 1 were uniformly low. The values for the highest score of 5 were more varied; these ranged from 2.1% for Lisbon to 14.2% for Coimbra. See Table 4.7 for all absolute numbers and percentages and Figures 4.2a to 4.2f for bar charts of the score distributions.

Score	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	0	0	0	0	0	0	0	0	1	1.2	2	2.7
2	7.5	12.3	9	16.4	9	13.4	16.5	23.6	15.5	18.9	11.5	15.8
3	23	37.7	25	45.5	31	46.3	35.5	50.7	49	59.8	38	52.1
4	25	41.0	17.5	31.8	17.5	26.1	16.5	23.6	11.5	14.0	16.5	22.6
5	5.5	9.0	3.5	6.4	9.5	14.2	1.5	2.1	5	6.1	5	6.8
Total	61	100.0	55	100.0	67	100.0	70	100.0	82	100.0	73	100.0

Table 4.7. Number of males with each supraorbital margin score, by collection

4.2.3 Mastoid Process

For the mastoid process, with sexes pooled (Table A2.4, Appendix 2), the significant differences in score distribution and median were between Grant and the other collections, due to the low number of Grant females. No significant differences lie between any of the other collections and each other. When females were considered alone (Table 4.8), after the Holm-Bonferroni correction was applied, no significant differences were found. For males only (Table 4.9), significant differences in both score distribution and central tendency were found between the Dart Collection compared to Grant and Lisbon, as well as between Pretoria and Grant. Significant differences in central tendency were also found between Coimbra and Dart and between Pretoria and Coimbra and between Pretoria and Lisbon.

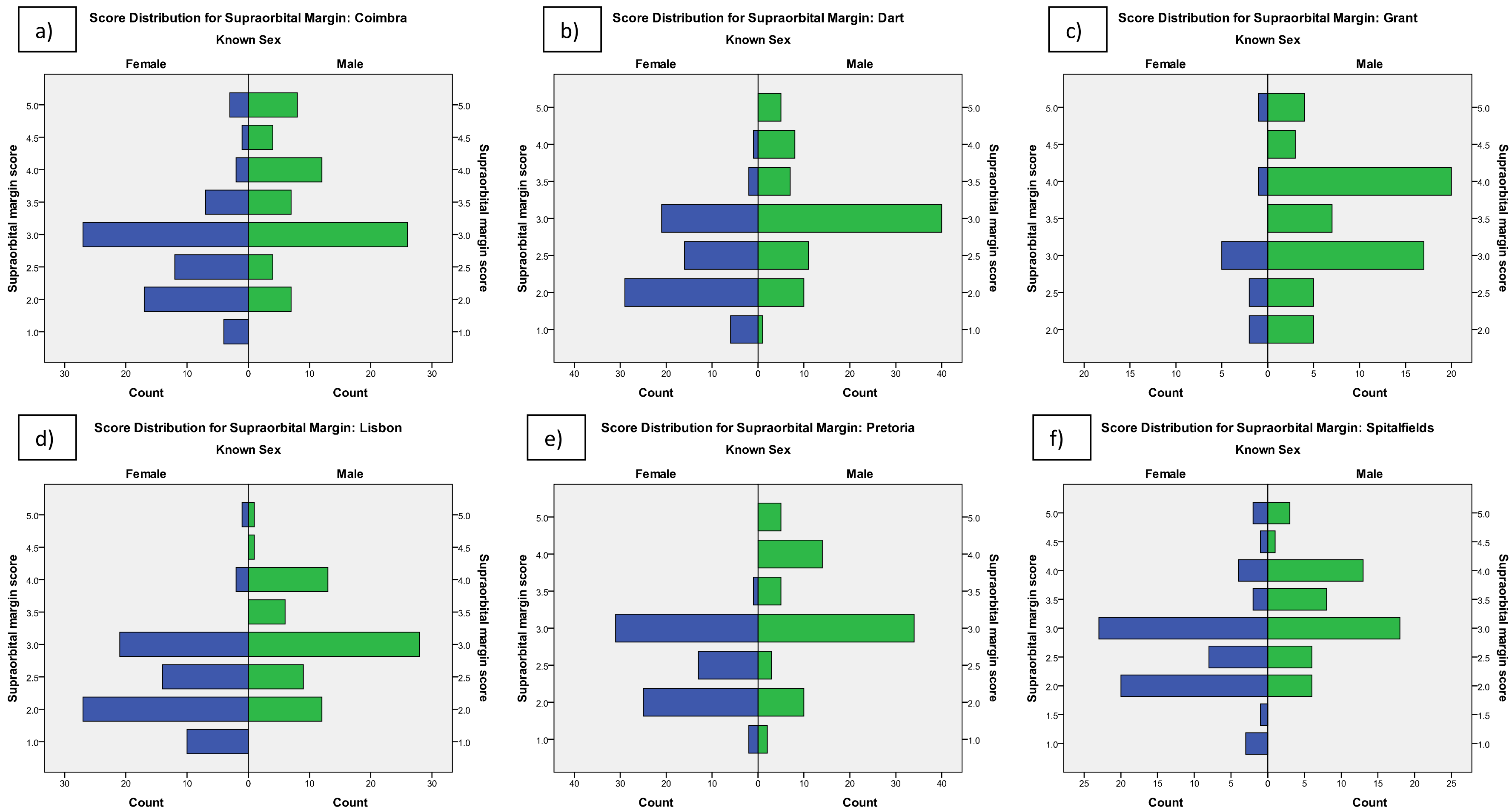
	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.996	.919	.636	.477	.280	.022	1.000	.822	.280	.438
Dart			.936	.399	.045	.020	.981	.720	.456	.554
Grant					.095	.038	.842	.605	.557	.402
Lisbon							.107	.015	.237	.287
Pretoria									.498	.375

Table 4.8. Mastoid process, females only: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.025	**.002**	.946	.208	1.000	.897	.008	**.001**	.999	.675
Dart			**.002**	**.000**	**.008**	**.001**	1.000	.782	.110	.025
Grant					.437	.220	**.000**	**.000**	.857	.116
Lisbon							.006	**.000**	.928	.572
Pretoria									.041	.017

Table 4.9. Mastoid process, males only: two-sample K-S and MWU results

When the distribution of female mastoid process scores against the known age distribution was examined, there were no age-related trends. Scores were approximately evenly distributed among the age groups – neither high nor low scores were reflective of older or younger ages. There were also no age-related trends in numbers of correctly-sexed females. In terms of total percentages of correctly-sexed females, Lisbon fared best, with 74.3% of females correctly identified using the mastoid process; Grant fared worst, with 36.4% correct. The other collections had similar results overall – Spitalfields females were correctly identified 58.6% of the time, Coimbra females, 58.3%, Dart females, 58.1% and Pretoria females, 55.6%. Full details can be found in Table A2.20 in Appendix 2.



Figures 4.2a to 4.2f. Supraorbital margin score distribution bar charts for each collection, separated by sex

By examining the score distributions, the locations for the differences in central tendency for Lisbon females compared to the other collections became apparent – there was a higher proportion of Lisbon females with scores of 2 and a lower proportion of scores of 3 compared to other collections, resulting in Lisbon's lower median. Compared to Dart, Lisbon had higher proportions of scores of 1 and 2 and lower proportions of scores of 3 and 4. For all collections except Grant, proportions of scores of 2 were the highest; scores of 3 were the next most common. For Grant, this pattern was reversed; scores of 3 were more common than scores of 2. No females had scores of 5, and scores of 4 were low for all collections. Scores of 1 occurred more often, ranging from 9.1% (Grant) to 25.8% (Spitalfields). It is interesting that Spitalfields had the highest proportions of both scores of 1 and 4 – morphological variation seems to be more widely ranging than for the other collections, where scores tended to concentrate more around 2 and 3. Table 4.10 and Figures 4.3a to 4.3f show the score distributions.

Score	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	1	9.1	16.5	25.8	8	11.1	14	18.4	9.5	12.8	10	13.9
2	4	36.4	21	32.8	34	47.2	42.5	55.9	33.5	45.3	30	41.7
3	5.5	50.0	19	29.7	23	31.9	16.5	21.7	25.5	34.5	24	33.3
4	0.5	4.5	7.5	11.7	7	9.7	3	3.9	5.5	7.4	8	11.1
5	0	0	0	0	0	0	0	0	0	0	0	0
Total	11	100.0	64	100.0	72	100.0	76	100.0	74	100.0	72	100.0

Table 4.10. Number of females with each mastoid process score, by collection

For males only, no clear age-related trends in scoring were observed. For instance, there was one male with a score of 1 from Coimbra, belonging to the 30 to 39 known-age group, and five males with scores of 2 in the lowest age groups (three in the 20 to 29 group and two in the 30 to 39 group); however, there were also two 60 to 69 year old males with scores of 2, and one 70 to 79 year old with a score of 2 to 3. None of the evidence seems to clearly support younger males tending to have lower scores (that is, the typically "female" scores). The percentages of correctly-sexed males may vary slightly with age – the youngest age group did not fare particularly well in any collection except Lisbon, ranging from 0.0% correct for Grant to 41.7% for Spitalfields (compared to 75% for Lisbon). However, low percentages of correctly-sexed males were also found in other age groups; there is no clear pattern of increasing or decreasing proportions of correctly-sexed males with age. When total percentage of correctly-sexed males was considered, the South African collections had the lowest numbers, resulting in only 26.0% and 28.0% sexed correctly, for Pretoria and Dart, respectively. The other collections had values of correctly-sexed males using only the mastoid process that were more similar to each other than to Dart and Pretoria – 48.2% for Spitalfields, 51.5% for Coimbra, 53.6% for Lisbon, and 58.9% for Grant. For full details, see Table A2.21 in Appendix 2.

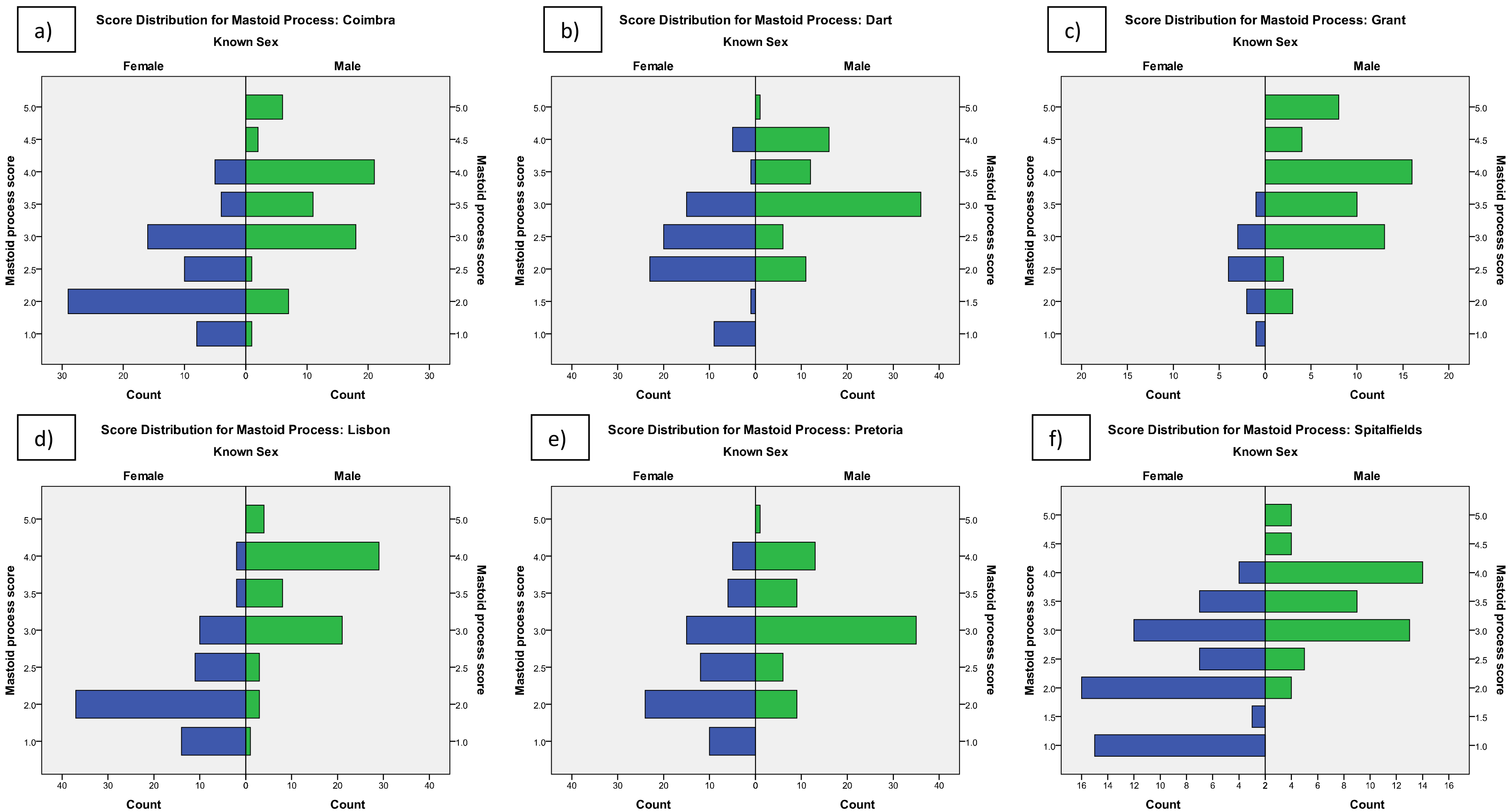
When the score distributions were examined in terms of proportions of males with each score, the differences between the South African collections and the other collections were located. Compared to Grant, Spitalfields, Coimbra and Lisbon, the Dart and Pretoria males tended to have lower scores, and had the largest proportions of scores of 3. The other collections had lower proportions of scores of 3. Dart and Pretoria also had higher percentages of males with scores of 2. The South African collections had correspondingly lower proportions of males with scores of 4 and 5 compared to the other collections. Scores of 1 either did not occur or were uncommon in all collections. Scores of 4 were the most common for Grant, Spitalfields, Coimbra and Lisbon, and scores of 3 were the next most common; the reverse was true of Dart and Coimbra. Scores of 5 were not hugely common for any collection, ranging from only one male each for Dart and Pretoria with scores of 5 (1.2% and 1.4%, respectively) to 17.9% of Grant males with scores of 5 for the mastoid process. Table 4.11 and Figures 4.3a to 4.3f show these values.

	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
Score	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	0	0	2	3.6	1	1.5	1	1.4	0	0	0	0
2	4	7.1	6.5	11.8	7.5	11.2	4.5	6.5	14	17.1	12	16.4
3	19	33.9	20	36.4	24	35.8	26.5	38.4	45	54.9	42	57.5
4	23	41.1	20.5	37.3	27.5	41.0	33	47.8	22	26.8	18	24.7
5	10	17.9	6	10.9	7	10.4	4	5.8	1	1.2	1	1.4
Total	56	100.0	55	100.0	67	100.0	69	100.0	82	100.0	73	100.0

Table 4.11. Number of males with each mastoid process score, by collection

4.2.4 Nuchal Crest

When the sexes were pooled for nuchal crest score distribution and median testing, statistically significant differences were found between the Grant Collection and every other collection. A significant difference in central tendency was found between Coimbra and Pretoria, Coimbra and Lisbon, Dart and Lisbon, and Dart and Pretoria. Table 4.12 provides all p-values.



Figures 4.3a to 4.3f. Mastoid process score distribution bar charts for each collection, separated by sex

For females only (Table 4.13), statistically significant differences in central tendency and score distribution were found between Coimbra and Dart. Significant differences in median only were found between Spitalfields and Pretoria, Coimbra and Pretoria, and Lisbon and Coimbra. Differences in distribution between Coimbra and Lisbon approached significance.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.038	.328	**.000**	**.000**	.026	**.001**	.073	**.006**	.736	.355
Dart			**.000**	**.000**	.020	**.003**	.136	**.026**	.262	.910
Grant					**.001**	**.000**	**.000**	**.000**	**.000**	**.000**
Lisbon							.996	.492	.052	.036
Pretoria									.303	.101

Table 4.12. Nuchal crest, sexes pooled: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	**.002**	**.004**	.049	.006	.011	**.001**	.005	**.000**	.300	.187
Dart			.493	.112	.701	.388	.079	.021	.090	.112
Grant					.954	.303	.989	.848	.350	.038
Lisbon							.164	.107	.241	.049
Pretoria									.044	**.003**

Table 4.13. Nuchal crest, females only: two-sample K-S and MWU results

Examining the score distributions by age group was not particularly revealing. The only collection for which the highest nuchal crest scores were strictly the domain of the older groups is the Dart Collection; the four individuals with scores of 4 were spread between the 70 to 79, 80 to 89 and 90 to 99 age groups. Females from other collections with morphologically male scores were spread throughout the age distribution. When percentages of correctly-identified females were inspected, no clear patterns of increasing or decreasing proportions with age were seen. In terms of the total percentage of correctly-identified females, the nuchal crest was most successful for the Coimbra Collection, at 75.3%, followed closely by the Spitalfields females, at 72.0%. Dart Collection females were successfully identified 67.6% of the time, while Lisbon females were correctly identified in only 56.0% of cases. Grant and Pretoria fared the worst, with 50.0% and 48.6%, respectively. Full details can be found in Table A2.22 of Appendix 2.

The reasons for the differences in distribution between Coimbra and all of the other collections except Spitalfields can be seen in the percent score distributions for each sample, presented in Table 4.14. Coimbra was the only collection in which the majority of females scored 1 for nuchal crest morphology. Spitalfields females were closest in terms of proportions of females with this score. However, for Coimbra, scores of 2, 3 and 4 were accordingly lower, decreasing in proportion with increasing score, but the distribution included one female with a score of 5. For all of the other collections, including Spitalfields, the majority of females had scores of 2. The morphologically male scores, 4 and 5, were the least common, although proportions of scores of 4 ranged from 21.2% for Pretoria to 4.8% for Coimbra. Coimbra had the most widely-spread values, whilst also having a rate of sexual dimorphism skewed towards the lowest, most female, morphology. Dart and Lisbon scores clustered largely around 2 and 3, while those of Spitalfields clustered around 1 and 2. Score distributions for Grant and Pretoria were more evenly spread between 2, 3 and 4. Figures 4.4a to 4.4f provide bar charts of the score distributions.

Score	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	0.5	4.2	17.5	29.7	31	42.5	12	16.0	9	12.2	10	13.7
2	5.5	45.8	25	42.4	24	32.9	30	40.0	41	55.4	25	34.2
3	3.5	29.2	9.5	16.1	13.5	18.5	27.5	36.7	20	27.0	22.5	30.8
4	2.5	20.8	7	11.9	3.5	4.8	5.5	7.3	4	5.4	15.5	21.2
5	0	0	0	0	1	1.4	0	0	0	0	0	0
Total	12	100.0	59	100.0	73	100.0	75	100.0	74	100.0	73	100.0

Table 4.14. Number of females with each nuchal crest score, by collection

When males were considered alone (Table 4.15), both the K-S and MWU tests found significant differences between Dart and Grant, Dart and Lisbon, and Grant and Pretoria. Further differences in central tendency were found between Coimbra and Grant, Coimbra and Lisbon, and Lisbon and Pretoria.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.623	.139	.008	**.000**	.043	**.002**	1.000	.710	.749	.363
Dart			**.000**	**.000**	**.000**	**.000**	.847	.319	.183	.041
Grant					.196	.173	**.001**	**.000**	.064	.019
Lisbon							.005	**.001**	.221	.126
Pretoria									.940	.254

Table 4.15. Nuchal crest, males only: two-sample K-S and MWU results

There are again no clear age-related patterns in male score distributions for the nuchal crest. Scores from either end of the spectrum (e.g. 1 and 5) generally occurred in any age group. However, the four Dart males with scores of 5 were aged 60 to 69 and older, while the four Pretoria males with scores of 5 were all in the 80 to 89 age group. It is possible that a slight trend is occurring – perhaps older males were more likely to exhibit the most extreme nuchal crest morphology (scores of 5); however, this is less evident for the non-South African collections. The results showing percentages of correctly-identified males by age group do not help to clarify the picture. High or low percentages occurred with no particular pattern of increase or decrease with age. The South African collections fared the worst in total percentages of correctly-identified males, at 29.9% and 38.4% for Dart and Pretoria, respectively. Coimbra followed closely with only 39.6% of males correctly sexed. Spitalfields males were sexed correctly in 48.0% of cases, while Lisbon males were correctly identified 60.7% of the time. The nuchal crest was most successful for sexing Grant Collection males, at 66.4%. Full details are in Table A2.23, Appendix 2.

The score distributions by percentage of each sample are presented in Table 4.16. The K-S and MWU tests showed significant differences between both Lisbon and Grant compared to Dart, and between Grant and Pretoria, but no significant differences existed between Lisbon and Grant or any other significant differences in score distribution. Central tendency differences also exist between Coimbra and Grant and Coimbra and Lisbon. Lisbon and Grant shared the highest proportion of males with scores of 4 (and lower proportions of scores of 3), while Coimbra, Dart and Pretoria shared the highest proportion of males with scores of 3 (and lower proportions of scores of 4). Grant and Lisbon also had much smaller proportions of males with scores of 2 compared to all other collections. For Spitalfields males, the highest proportion of males had scores of 4, like Lisbon and Grant, but scores of 3 were nearly as common; also, low scores of 2 were more common than for either Grant or Lisbon. This explains the significant difference in central tendency between Spitalfields and Grant, while the high proportions of scores of 4 and correspondingly lower scores of 3 and 2 explain why Grant and Lisbon had significantly different score distributions compared to Coimbra, Dart and Pretoria. Scores of 5 were not common for Spitalfields, Coimbra, Lisbon, Dart or Pretoria, but for Grant were as high as 18.0%. No scores of 1 occurred in males from any collection sampled here. Table 4.16 and Figures 4.4a to 4.4f show all details.

	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
Score	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	0	0	0	0	0	0	0	0	0	0	0	0
2	4	6.6	8	16.0	10	14.9	1	1.4	19	23.2	14.5	19.9
3	16.5	27.0	18	36.0	30.5	45.5	26.5	37.9	38.5	47.0	30.5	41.8
4	29.5	48.4	20	40.0	24.5	36.6	38.5	55.0	20.5	25.0	24	32.9
5	11	18.0	4	8.0	2	3.0	4	5.7	4	4.9	4	5.5
Total	61	100.0	50	100.0	67	100.0	70	100.0	82	100.0	73	100.0

Table 4.16. Number of males with each nuchal crest score, by collection

4.2.5 Mental Eminence

For the mental eminence, when sexes were pooled, significant differences in median were found between Dart and Grant, and Grant and Spitalfields. Table 4.17 contains all the p-values, below.

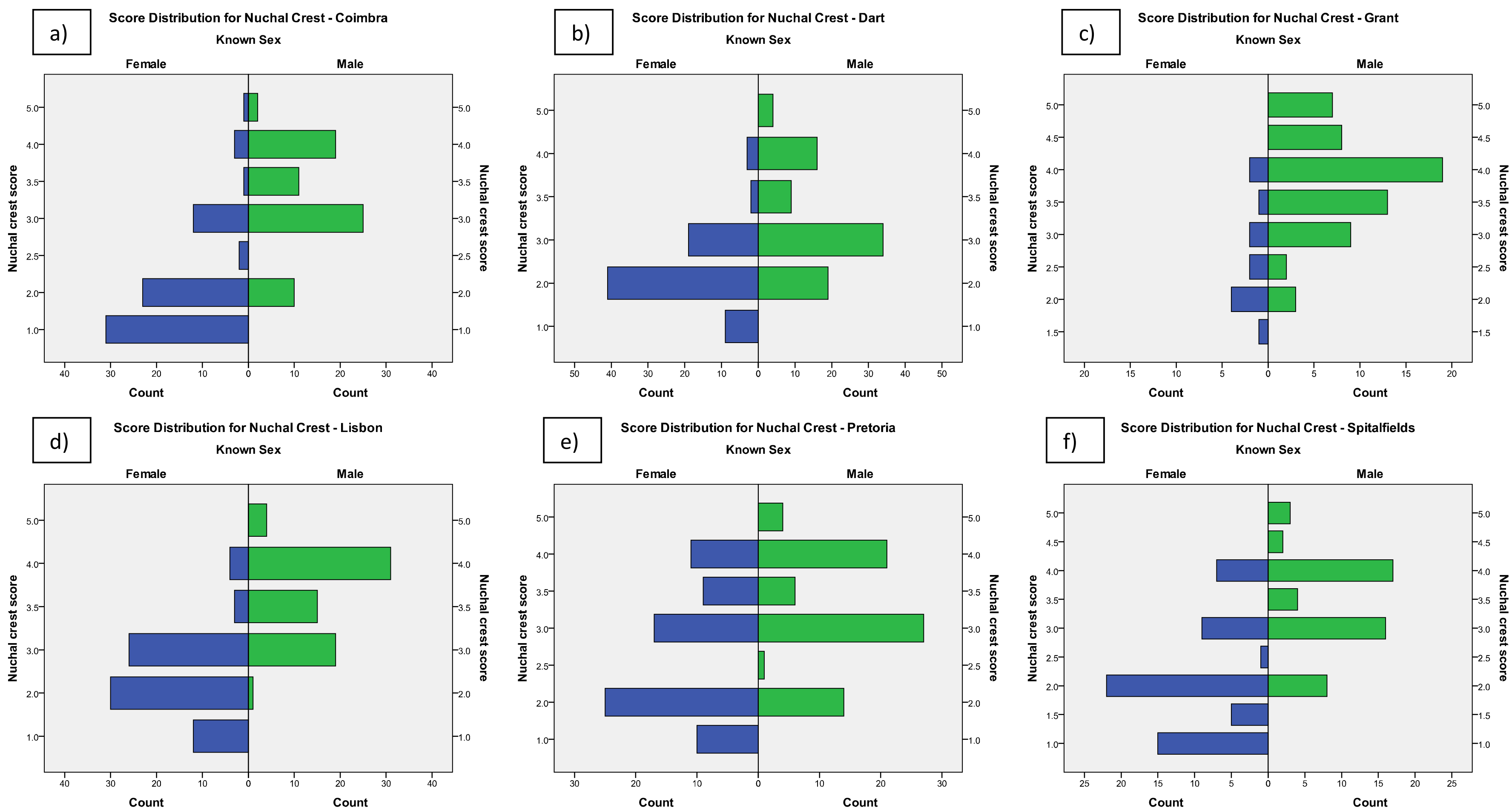
	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.953	.555	.051	.006	.809	.251	.997	.675	.681	.402
Dart			.014	**.001**	.647	.063	.887	.267	.370	.634
Grant					.166	.051	.087	.009	.025	**.002**
Lisbon							.538	.441	.183	.052
Pretoria									.271	.224

Table 4.17. Mental eminence, sexes pooled: two-sample K-S and MWU results

No statistically significant differences in score distribution were found between collections for either females or males alone. For females alone (Table A2.5, Appendix 2), differences in median between Lisbon and Spitalfields and Pretoria and Spitalfields neared significance. Similarly, for males alone (Table A2.6, Appendix 2), only Dart and Lisbon had a difference in median that approached significance.

When scores were analysed by age group, no age-related patterns were observed. Neither higher, morphologically “male” scores, nor lower, morphologically “female” scores were the domain of either younger or older ages. Scores were distributed amongst the age groups fairly evenly. When looking at the percentages of correctly-identified females by age group, there was again no clear age-related pattern. The mental eminence alone did quite well at identifying sex – as high as 77.3% of females were correctly identified, for the Dart Collection, followed by 75.0% for Spitalfields females and 71.2% for Coimbra females. Grant, Lisbon, and Pretoria females had similar proportions of correct sex identification, at 64.3%, 65.1% and 64.3%, respectively. Table A2.24 in Appendix 2 has full details.

No statistically significant differences in female score distribution between collections were found for the mental eminence. For all collections, scores of 2 were the most common, followed by scores of 3 (considered to be unable to determine sex). The differences in central tendency between Spitalfields and Lisbon and Spitalfields and Pretoria stemmed from the fact that Spitalfields had a higher proportion of scores of 1 and a lower proportion of scores of 3 compared to Lisbon and Pretoria. Coimbra had a higher proportion of scores of 1 and lower proportion of



Figures 4.4a to 4.4f. Nuchal crest score distribution bar charts for each collection, separated by sex

scores of 3 compared to Lisbon and Pretoria, while the bulk of Dart females had scores of 2, so that the central tendencies for these two collections were not significantly different from any of the others. The Grant Collection's seven individuals all scored either 2 (the majority) or 3. Across all collections, there was only one score of 5, from a Lisbon female. Scores of 4 were low across all collections. Details are in Table 4.18 and Figures 4.5a to 4.5f, below.

Score	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	0	0	10	16.7	9.5	13.0	5	6.8	2.5	3.3	5	7.9
2	4.5	64.3	35	58.3	42.5	58.2	42.5	58.2	55.5	74.0	35.5	56.3
3	2.5	35.7	13	21.7	19.5	26.7	22.5	30.8	16	21.3	21.5	34.1
4	0	0	2	3.3	1.5	2.1	2	2.7	1	1.3	1	1.6
5	0	0	0	0	0	0	1	1.4	0	0	0	0
Total	7	100.0	60	100.0	73	100.0	73	100.0	75	100.0	63	100.0

Table 4.18. Number of females with each mental eminence score, by collection

No age-related patterns in male score distribution for mental eminence were apparent. Young and old individuals were equally likely to have a higher or lower score (although the only male score of 1 belongs to a Spitalfields male from the 30 to 39 age group). Scores of 5, more common in Coimbra and Lisbon males, were spread over the age distributions. The percentages of correctly-identified males also show no age-related patterns. Total percentages of correctly-identified males were low for all collections. Dart males had the lowest percentage of correct sex identification, at 23.8%, followed by Pretoria, at 29.2%. Spitalfields and Coimbra had similar proportions of correctly identified males, at 31.1% and 31.3%, respectively. Grant Collection males were correctly sexed in 34.3% of cases, while Lisbon fared best, its males correctly sexed in 39.1% of cases. Full details are in Table A2.25 of Appendix 2.

The score distributions show that the most common score across all collections sampled was 3. Scores of 4 (considered morphologically "male") were the second most common. Scores of 2 (considered morphologically "female") were the next most common score. Scores of 1 were uncommon; only one Spitalfields male had a score of 1. Scores of 5 occurred, but not frequently, from a low of 0.8% for Pretoria to a high of 10.1% for Lisbon. The only significant difference was in central tendency between Lisbon and Dart, which was the result of Lisbon's higher proportions of scores of 4 and 5, and half as many scores of 1 as Dart. All collections had scores that are fairly well spread across the possible score range. Percentages and absolute numbers are in Table 4.19, and bar charts are in Figures 4.5a to 4.5f.

Score	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	0	0	1	1.9	0	0	0	0	0	0	0	0
2	4.5	12.9	6.5	12.3	7	10.4	6	8.7	12.5	15.6	10	15.4
3	18.5	52.9	29	54.7	39	58.2	36	52.2	48.5	60.6	36	55.4
4	10.5	30.0	16	30.2	15.5	23.1	20	29.0	18	22.5	18.5	28.5
5	1.5	4.3	0.5	0.9	5.5	8.2	7	10.1	1	1.3	0.5	0.8
Total	35	100.0	53	100.0	67	100.0	69	100.0	80	100.0	65	100.0

Table 4.19. Number of males with each mental eminence score, by collection

4.3 Sex Determination Using Specific Features of the Pelvis

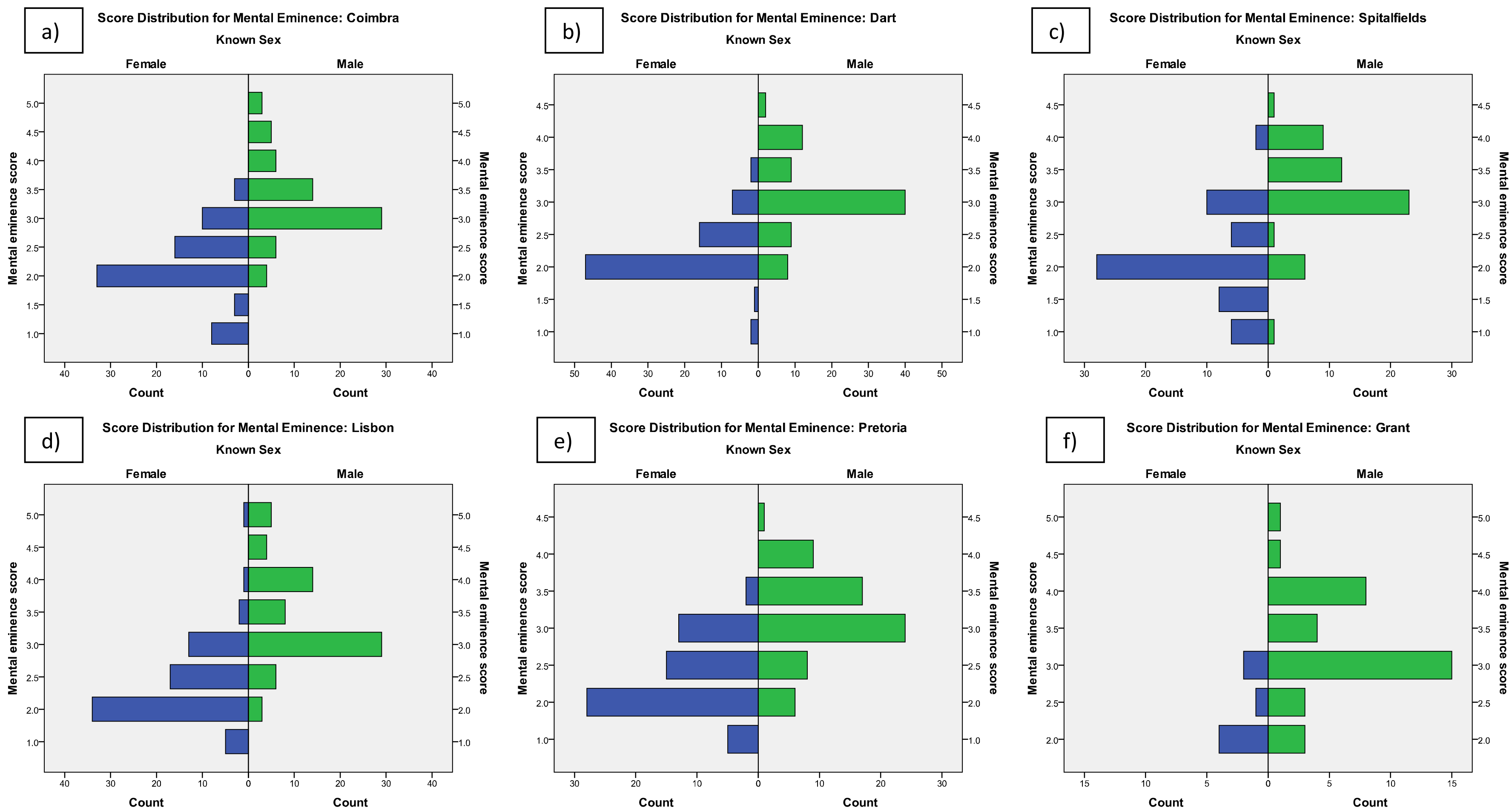
4.3.1 Sciatic Notch

The greater sciatic notch scores were first examined with the sexes pooled (Table A2.7, Appendix 2). Significant differences in both score distribution and median were found between Grant and every other collection. The difference in score distribution between Dart and Lisbon approached significance. When females only were considered, no significant differences, in either score distribution or central tendency, were observed (Table A2.8, Appendix 2). However, when males only were considered, a significant difference in score distribution was found between Grant and Lisbon, and a significant difference in central tendency was found between Dart and Lisbon (Table 4.20, below).

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.190	.161	.077	.276	.805	.071	1.000	.946	.489	.663
Dart			.998	.792	.004	**.003**	.381	.182	.995	.446
Grant					**.001**	.008	.167	.301	.998	.622
Lisbon							.459	.073	.029	.064
Pretoria									.764	.692

Table 4.20. Sciatic notch, males only: two-sample K-S and MWU results

Observation of the female score distribution by age group did not reveal any age-related patterning. Lower and higher scores occurred at all ages. The distribution of proportions of correctly-identified females by age group also did not reveal any age-related trends. In terms of total percentages of correctly-identified females, the sciatic notch was least successful for the Coimbra Collection, at 51.4%, followed by 52.1% for Dart females and 53.3% for Lisbon females. Slightly better results were found for the Pretoria females, at 57.6%; the sciatic notch was most successful for Spitalfields and Grant females, at 61.5% and 63.9%, respectively. Full details can be found in Table A2.26, in Appendix 2.



Figures 4.5a to 4.5f. Mental eminence score distribution bar charts for each collection, separated by sex

While no significant differences were found between female score distributions of the six collections, it is useful to examine the proportions of scores nonetheless, found in Table 4.21 and Figures 4.6a to 4.6f, below. The most common score for all collections except Dart was 2; for Dart, scores of 3 were slightly more common. Scores of 3 were the next most common (except for Dart). For Coimbra, Dart and Pretoria, the difference between females with scores of 2 and females with scores of 3 was small, only two or fewer individuals. For Spitalfields, Coimbra, Dart and Pretoria, scores of 1 were more common than scores of 4; for Lisbon, scores of 4 were slightly more common than scores of 1. Grant had equal numbers of females with scores of 1 and 4. Proportions of scores of 1 ranged from 7.5% for Coimbra to 23.1% for Spitalfields; scores of 4 ranged in percentage from 3.1% for Spitalfields to 13.7% for Dart. Scores of 5 were extremely uncommon.

Score	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	2.5	13.9	15	23.1	5.5	7.5	6	7.9	14.5	19.9	13	18.1
2	9	50.0	25	38.5	32	43.8	34.5	45.4	23.5	32.2	28.5	39.6
3	4	22.2	21	32.3	30.5	41.8	27	35.5	24	32.9	27	37.5
4	2.5	13.9	2	3.1	5	6.8	8.5	11.2	10	13.7	3.5	4.9
5	0		2	3.1	0		0		1	1.4	0	
Total	18	100.0	65	100.0	73	100.0	76	100.0	73	100.0	72	100.0

Table 4.21. Number of females with each sciatic notch score, by collection

As with the females, the male score distribution by age group did not display any age-related trends. The percentages of correctly-identified males by age group also did not show any clear age-related patterns. The sciatic notch was more successful at determining the sex of males compared to the females. The lowest total percentage of correctly-identified males was 72.9% for Lisbon, followed by 74.1% for Spitalfields – these least-successful results were higher than the most-successful female results (63.9% for Grant females). The sciatic notch successfully identified 77.7% of Grant males, 78.0% of Dart males, and 78.4% of Pretoria males. It was most successful for Coimbra males, correctly identifying 81.3% of them. Full details can be found in Table A2.27 of Appendix 2.

By looking at the percentages of males with each score per collection, it can be seen that the differences between Lisbon compared to Dart and Grant lie in the proportion of scores of 4 and 5. The distributions of 1, 2, and 3 between these collections (indeed, any of the collections) were similar – only Grant, Spitalfields and Pretoria had scores of 1, while scores of 2 were also few or non-existent. However, Lisbon had a high percentage of scores of 4 and a relatively low percentage of scores of 5 compared to Grant and Dart. Indeed, the most common score for

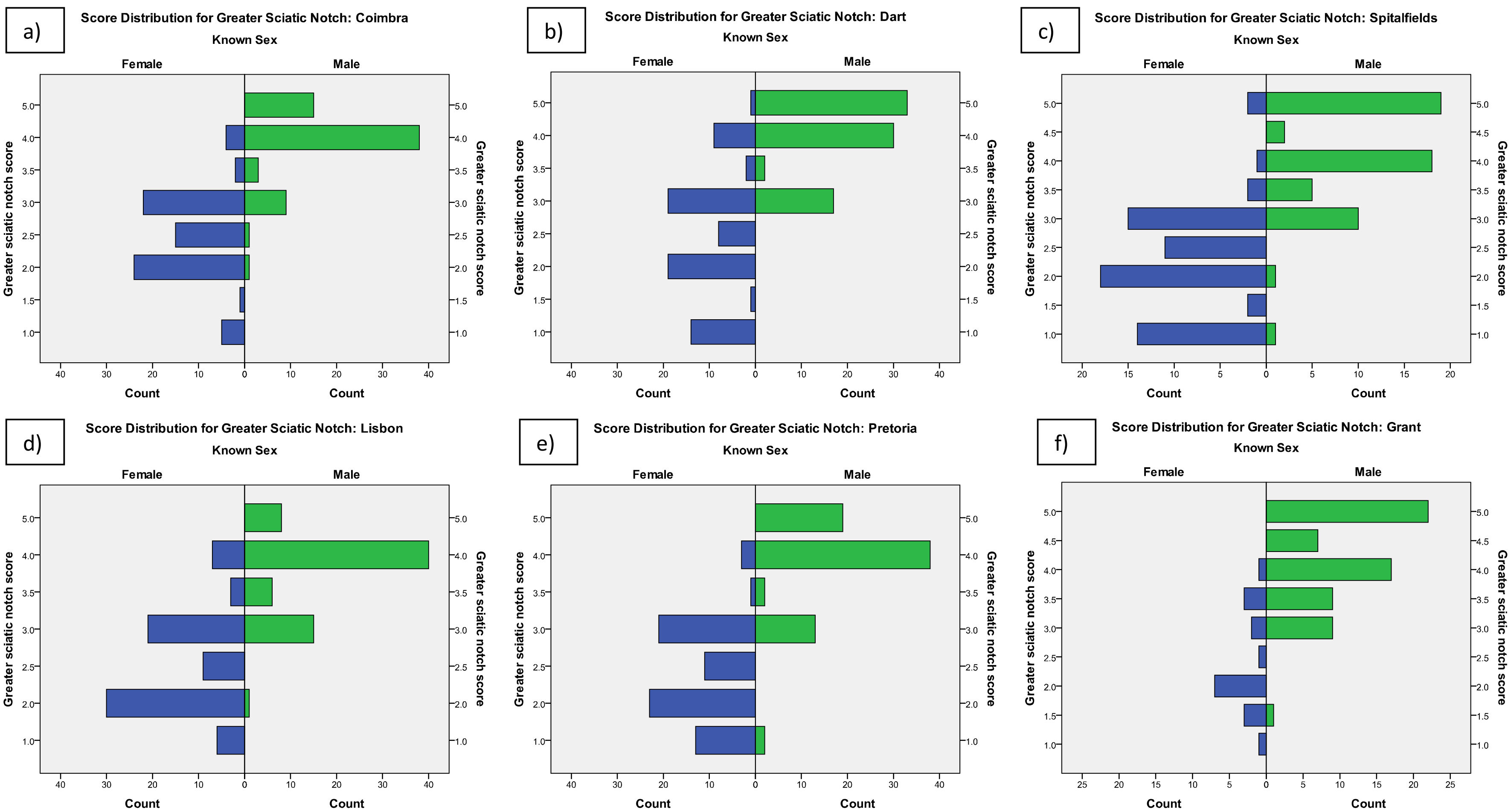
Lisbon was 4, while for both Dart and Grant, it was 5. Between Lisbon and Spitalfields, the difference again was largely in the proportion of scores of 4 and 5 – Lisbon again had a comparatively high proportion of scores of 4 and a low proportion of scores of 5 compared to Spitalfields. The slightly higher proportions of lower scores for Spitalfields males compared to Grant and Dart mean that Spitalfields' central tendency was slightly lower and not significantly different to that of Lisbon. In general, scores of 1 and 2 were very uncommon. Scores of 3 were more common. Scores of 4 were the most common score for Spitalfields, Coimbra, Lisbon and Pretoria. Scores of 5 were the most common score for Grant and Dart. The distributions for Grant, Spitalfields and Dart were spread more evenly across scores 3, 4, and 5, while the distributions of Coimbra, Lisbon and Pretoria were more highly concentrated (around scores of 4). Table 4.22 and Figures 4.6a to 4.6f show percentages and bar charts, respectively.

Score	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	0.5	0.8	1	1.8	0		0		0		2	2.7
2	0.5	0.8	1	1.8	1.5	2.2	1	1.4	0		0	
3	13.5	20.8	12.5	22.3	11	16.4	18	25.7	18	22.0	14	18.9
4	25	38.5	21.5	38.4	39.5	59.0	43	61.4	31	37.8	39	52.7
5	25.5	39.2	20	35.7	15	22.4	8	11.4	33	40.2	19	25.7
Total	65	100.0	56	100.0	67	100.0	70	100.0	82	100.0	74	100.0

Table 4.22. Number of males with each sciatic notch score, by collection

4.3.2 Ischiopubic Ramus Ridge

For the sexes together (see Table 4.23, below), statistically significant differences were found for the ischiopubic ramus between Grant and every other collection. A significant difference in score distribution (and approaching significance in median) was also found between Pretoria compared to Coimbra and Dart; differences in distribution and median between Pretoria and Lisbon approached significance. However, for females alone (Table 4.24), significant differences in score distribution and median were only found between Coimbra and Pretoria and Lisbon and Pretoria. No other significant differences were found. For males only (Table 4.25), significant differences were found with the K-S test and MWU test between Pretoria compared to Dart and Grant. Significant differences in median only were found between Grant compared to Coimbra and Lisbon.



Figures 4.6a to 4.6f. Sciatic notch score distribution bar charts for each collection, separated by sex

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.050	.673	**.000**	**.000**	.443	.887	**.001**	.007	.423	.444
Dart			**.000**	**.000**	.115	.763	**.002**	.011	.296	.369
Grant					**.000**	**.000**	**.000**	**.000**	**.000**	**.000**
Lisbon							.023	.009	.590	.469
Pretoria									.438	.123

Table 4.23. Ischiopubic ramus ridge, sexes pooled: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.092	.090	.987	.507	.803	.837	**.002**	**.000**	.515	.147
Dart			1.000	.670	.067	.096	.093	.059	.875	.737
Grant					.962	.538	.722	.071	1.000	.818
Lisbon							**.001**	**.000**	.412	.166
Pretoria									.347	.024

Table 4.24. Ischiopubic ramus ridge, females only: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.176	.070	.063	**.003**	.579	.491	.014	.034	1.000	.953
Dart			.627	.297	.084	.026	**.003**	**.001**	.378	.120
Grant					.028	**.001**	**.001**	**.000**	.166	.007
Lisbon							.575	.175	.993	.560
Pretoria									.154	.067

Table 4.25. Ischiopubic ramus ridge, males only: two-sample K-S and MWU results

The female score distributions by age group showed no age-related patterning; higher, morphologically “male” scores occurred (infrequently) in all age groups. When the percentages of correctly-identified females by age group were examined, no age-related patterns were revealed. In terms of total percentages of correctly identified females, Pretoria fared best at 91.5%, followed by 81.5% for Spitalfields females. Lisbon females were identified correctly in 79.7% of cases; for Grant, 75.0% of females were correctly sexed. For Dart females, 70.8% were correctly identified; the ischiopubic ramus ridge was least successful for Coimbra, where 68.6% of females were correctly identified. Full details are in Table A2.28 of Appendix 2.

When the score distributions as percentages of each sample were observed, the reasons for the significant differences in distribution and central tendency were seen. Pretoria females had significant differences in distribution and median compared to both Coimbra and Lisbon; in Table 4.26 and Figures 4.7a to 4.7f, below, the reason seems to be that Pretoria had a larger proportion of scores of 1, and smaller proportions of every other score. Scores of 2 were the second highest; for Pretoria, the proportion is 9.9%, while for Coimbra and Lisbon it is 18.6% and 31.3%, respectively. Proportions of higher scores were lower still. The difference in central tendency between Pretoria and Spitalfields again seems largely due to the higher proportion of scores of 1 in Pretoria females. No other significant differences were detected. For all collections, the most common score was 1, ranging from Pretoria's high of 81.7% to 48.4% for Lisbon females. Scores of 2 were the second most common for Pretoria, Lisbon and Spitalfields; scores of 3 were the second most common for Coimbra and Dart. For Grant, there were equal numbers of females with scores of 2 and 3. Scores of 4 and 5 were fairly uncommon; interestingly, scores of 4 were less common than scores of 5. Proportions of scores of 5 ranged from 0% for Pretoria to 13.0% for Spitalfields. Score distributions for Coimbra, Grant, and, to some extent, Spitalfields and Lisbon, were spread more evenly than the distributions of Pretoria and Dart, which were more highly peaked around scores of 1.

Score	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	n	% of *n*
1	10	62.5	35	64.8	35	50.0	31	48.4	51	70.8	58	81.7
2	2	12.5	9	16.7	13	18.6	20	31.3	0	0.0	7	9.9
3	2	12.5	3	5.6	14	20.0	4	6.3	13	18.1	2	2.8
4	0		0		5	7.1	3	4.7	0	0.0	4	5.6
5	2	12.50	7	13.0	3	4.3	6	9.4	8	11.1	0	
Total	16	100.0	54	100.0	70	100.0	64	100.0	72	100.0	71	100.0

Table 4.26. Number of females with each ischiopubic ramus score, by collection

For males, when the score distribution by age group was observed, no age-related trends could be seen. Higher, morphologically "male" scores were more common in general, but lower, morphologically "female" scores were possible in any age group. Similarly, no age-related trends were observed in the percentages of correctly-identified males by age group. In terms of total percentages of correctly-sexed males, Pretoria fared the worst at 67.6%, followed by 78.8% for Lisbon males. For Spitalfields, 81.8% of males were correctly identified, while 82.9% of Dart males and 84.6% of Coimbra males were correctly identified. The ischiopubic ramus ridge performed best for Grant males, of whom 95.4% were correctly identified. Table A2.29 in Appendix 2 has full details.

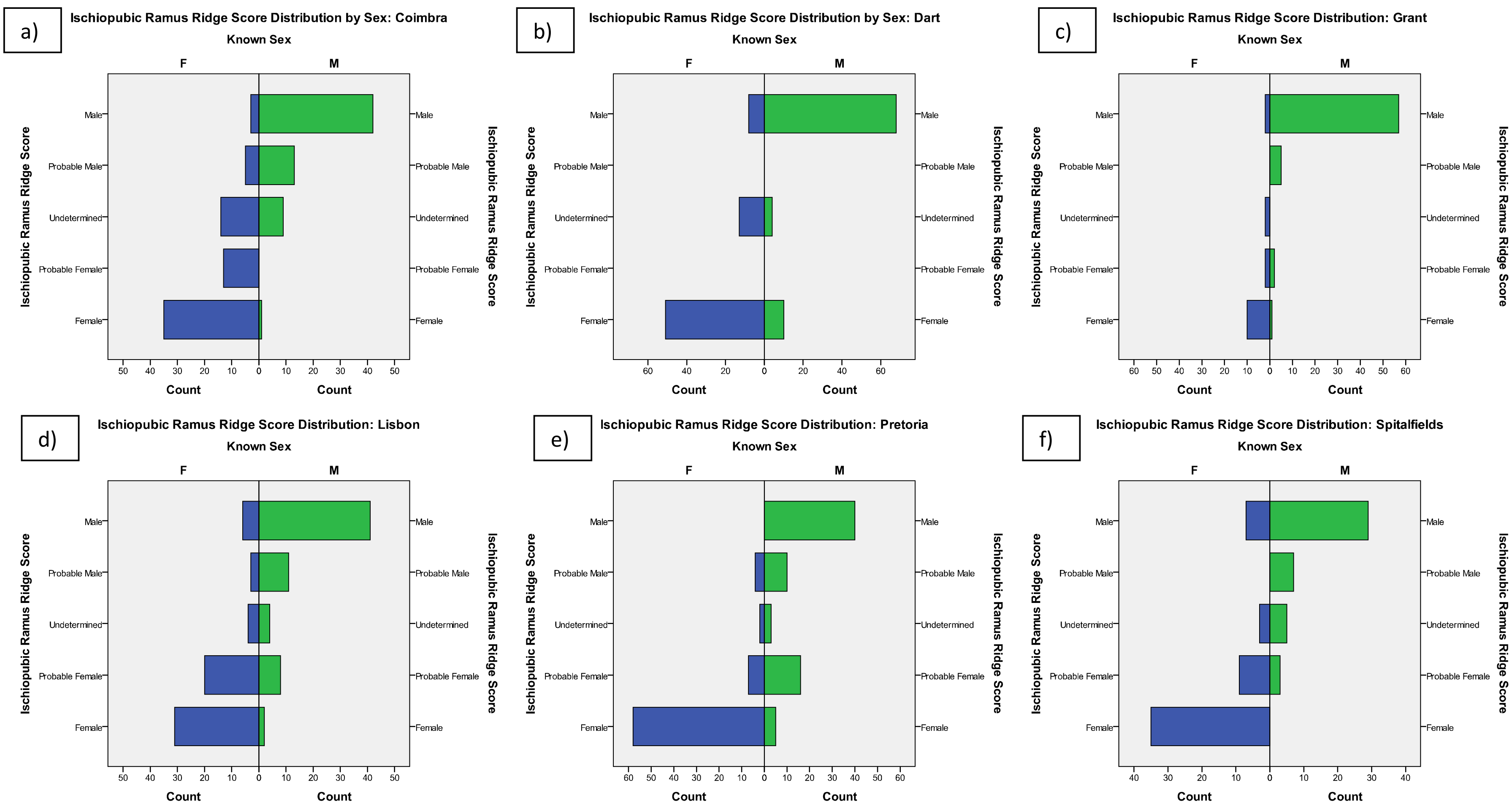
Examination of the distribution of the scores by proportion (in Table 4.27 and Figures 4.7a to 4.7f, below) showed the locations of the differences found by the K-S test and reasons for differences in central tendency. A significant difference in central tendency was found between Grant and Lisbon; the reason seems to be because scores for Grant were more highly peaked around scores of 5. Lisbon males also have at least double the proportion of every other score compared to Grant. Pretoria also had significant differences in distribution and median compared to Dart and Grant. The score distribution for Pretoria males appears to be bimodal; there were peaks at scores of 2 and 5. Neither Coimbra nor Grant were bimodal; low scores were uncommon for both, with scores of 4 and 5 being most common for Grant, and scores of 3, 4 and 5 being most common for Coimbra. Meanwhile, the score distribution for Dart also seems somewhat bimodal (with peaks at scores of 1 and 5). The differences between Pretoria and Dart lie more in the locations of the peaks, and the fact that Dart had no males with scores of 2 or 4, while there were males with every score from Pretoria. Grant also had a significantly different median from Coimbra and nearing significance compared to Spitalfields – this stems from the fact that the vast majority of Grant males had scores of 5 and few lower scores, while both Coimbra and Spitalfields had more males with scores of 3 and 4, resulting in the higher Grant median. Dart and Lisbon also had a difference nearing significance in median; here again, Dart's median was higher due to the majority of males having scores of 5. Lisbon also had higher proportions of males with scores of 2, 3 and 4. In general, the most common score was 5. The second most common score was 4 for Grant, Spitalfields, Coimbra and Lisbon males. However, for Dart, the second most common score was 1, while for Pretoria, it was 2. In all but Dart males, scores of 1 were uncommon, ranging from 0.0% for Spitalfields to 6.8% for Pretoria. Scores of 2 were uncommon for Grant, Spitalfields, Coimbra and Dart males.

	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
Score	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	1	1.5	0	0.0	1	1.5	2	3.0	10	12.2	5	6.8
2	2	3.1	3	6.8	0	0.0	8	12.1	0	0.0	16	21.6
3	0	0.0	5	11.4	9	13.8	4	6.1	4	4.9	3	4.1
4	5	7.7	7	15.9	13	20.0	11	16.7	0	0.0	10	13.5
5	57	87.7	29	65.9	42	64.6	41	62.1	68	82.9	40	54.1
Total	65	100.0	44	100.0	65	100.0	66	100.0	82	100.0	74	100.0

Table 4.27. Number of males with each ischiopubic ramus score, by collection

4.3.3 Subpubic Concavity

The K-S and MWU tests for the subpubic concavity for the sexes pooled revealed significant differences between the Grant Collection and every other collection. See Table A2.9, Appendix 2 for details. Interestingly, for both males and females alone, no significant differences in score distribution were found (see Tables A2.10 and A2.11, Appendix 2). For females alone, only the difference in median between Coimbra and Spitalfields approached significance. For males alone, significant differences in median were found between Dart and Spitalfields, and Lisbon and Spitalfields.



Figures 4.7a to 4.7f. Ischiopubic ramus ridge score distribution bar charts for each collection, separated by sex

No age-related trends in score distribution were observed in the female score data divided by age group. The few high, morphologically “male” scores that did appear were scattered throughout the age groups, regardless of the collection. As the scores were largely concentrated at the morphologically “female” end of the scale (scores of 1 and 2), the subpubic concavity performed quite well in correctly identifying females, regardless of age group. The total percentages of correctly identified females reflect the success of the subpubic concavity for sex determination in this study; 100.0% of Pretoria females were correctly identified, followed closely by 98.6% for Coimbra females and 98.2% for Spitalfields females. For Grant females, the subpubic concavity was successful in 94.4% of cases, followed by 93.1% for Dart females. The subpubic concavity was least successful for Lisbon females, where 92.8% were correctly identified. Full details are in Table A2.30, Appendix 2.

The only significant differences for females were in central tendency between Coimbra and Grant, Coimbra and Lisbon, and Coimbra and Spitalfields. Upon observation of the score distribution by percentage of each sample, it is revealed that the reason for the differences between Coimbra and Grant, Spitalfields, and Lisbon was that Coimbra had a higher proportion of scores of 1 than the other collections and accordingly lower proportions of other scores. Scores of 2 were relatively more common for Grant, Spitalfields and Lisbon females. A few scores of 4 and 5 appeared in Spitalfields and Lisbon females. In general, scores of 1 were the most common, accounting for the vast majority of all females sampled. All other scores were fairly rare, but scores of 2 were the next most common for Grant, Spitalfields and Pretoria. See Table 4.28 and Figures 4.8a to 4.8f.

Score	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	15	83.3	49	87.5	69	97.2	60	87.0	67	93.1	68	95.8
2	3	16.7	6	10.7	1	1.4	4	5.8	0	0.0	3	4.2
3	0	0.0	0	0.0	1	1.4	0	0.0	0	0.0	0	0.0
4	0	0.0	0	0.0	0	0.0	1	1.4	0	0.0	0	0.0
5	0	0.00	1	1.8	0	0.0	4	5.8	5	6.9	0	0.0
Total	18	100.0%	56	100.0%	71	100.0%	69	100.0%	72	100.0%	71	100.0%

Table 4.28. Number of females with each subpubic concavity score, by collection

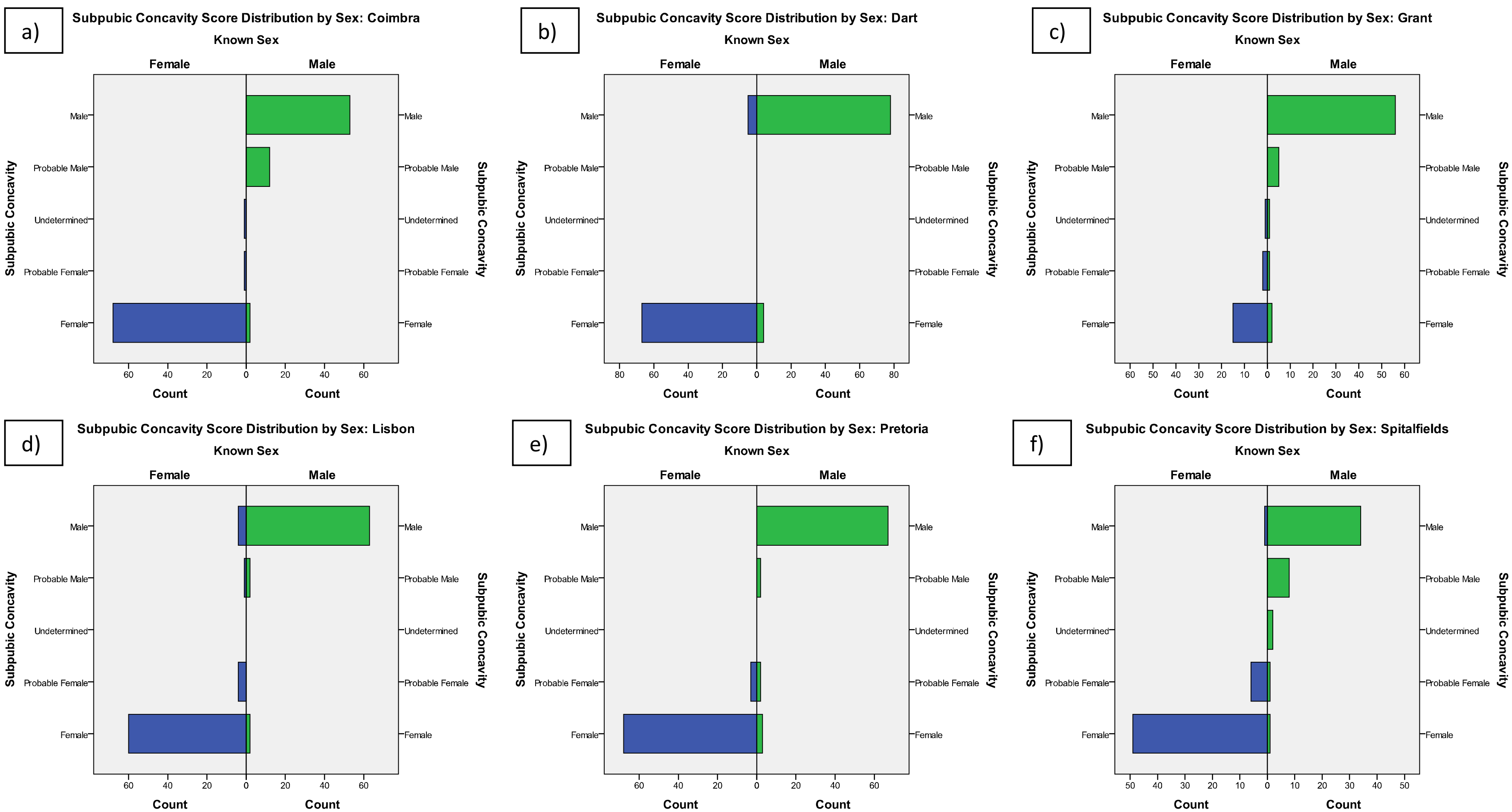
As with females, the males showed no age-related trends in score distribution by age group. The majority of males, across all collections, had morphologically “male” scores of 4 or 5, but where there were lower scores, they were not restricted to any particular age group. Because most males scored 4 or 5 (morphologically “male”), there does not seem to be any age-related patterning to the percentage of correctly-identified males by age group. The total percentages of correctly-identified males were accordingly high – 91.3% for Spitalfields, 93.2% for Pretoria, 93.8% for Grant, 95.1% for Dart, and 97.0% for both Coimbra and Lisbon. Full details are in Table A2.31, Appendix 2.

Significant differences in central tendency were found between Coimbra and Dart, Coimbra and Lisbon, and Spitalfields compared to Dart, Lisbon, and Pretoria. The reasons for such differences can be seen by examining the score distribution for each collection as percentages of the total sample (Table 4.29, below). Coimbra's central tendency was lower than that of either Dart or Lisbon because of a larger percentage of males with scores of 4. Accordingly, Coimbra's percentage of males with scores of 5 was also lower than that of Lisbon or Dart. None of these three collections had males with scores of 2 or 3, and all had low proportions of scores of 1; thus, the difference lies in the proportions of scores of 4 and 5. The central tendency for Spitalfields was significantly different from that of Lisbon, Dart and Pretoria for similar reasons – a comparatively lower proportion of scores of 5, and higher proportions of scores of 4 and 3. Scores of 1 and 2 were low for Spitalfields, Lisbon, Dart and Pretoria. In general, scores of 5 were the most common by far. Scores of 4 were the second most common for Grant, Spitalfields and Coimbra. Scores of 1 were the second most common for Dart and Pretoria males. Figures 4.8a to 4.8f provide bar charts of the score distributions.

Score	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	2	3.1	1	2.2	2	3.0	2	3.0	4	4.9	3	4.1
2	1	1.5	1	2.2	0	0.0	0	0.0	0	0.0	2	2.7
3	1	1.5	2	4.3	0	0.0	0	0.0	0	0.0	0	0.0
4	5	7.7	8	17.4	12	17.9	2	3.0	0	0.0	2	2.7
5	56	86.15	34	73.9	53	79.1	63	94.0	78	95.1	67	90.5
Total	65	100.0	46	100.0	67	100.0	67	100.0	82	100.0	74	100.0

Table 4.29. Number of males with each subpubic concavity score, by collection

There were again significant differences between the collections when the sexes were pooled that were not explained by variation in either sex on its own. Significant differences in the distributions of Spitalfields and Dart were found because the vast majority of Dart females had the most "female" morphology (93.1% of females had the lowest score of 1), and the vast majority of Dart males had the most "male" morphology (95.1% of males had the highest score of 5). Conversely, the Spitalfields scores were more evenly distributed, with some females scoring 2 and some males scoring 4. The distribution was still distinctly bimodal, but not in the extreme as it was for Dart.



Figures 4.8a to 4.8f. Subpubic concavity score distribution bar charts for each collection, separated by sex

4.3.4 Ventral Arc

When the sexes were pooled for the ventral arc, the K-S and MWU tests showed significant differences between the Grant Collection and every other collection (see Table A2.12, Appendix 2). Neither males alone nor females alone showed any significant differences in score distribution (see Tables A2.13 and A2.14, Appendix 2). For females only, differences in central tendency between Coimbra and Dart, Coimbra and Pretoria, Dart and Grant, Grant and Pretoria, and Pretoria and Spitalfields neared significance. For males only, no significant differences in central tendency were found.

No age-related trends were observed in the female score distribution by age group. Where higher, morphologically "male" scores were present, they seem to occur at random, and not in specific age ranges. No age-related trends were observed in terms of percentages of correctly-identified females by age group. Although the youngest and oldest age groups tended to have the lowest percentages of correctly-identified females, it is worth noting that these groups had lower absolute numbers of females; generally, these low percentages still represent only one or two incorrectly-identified females, the same absolute number found in other age groups. As such, it does not seem appropriate to consider the somewhat lower percentages of correct sex identification for the youngest and oldest groups as a trend. As for the other single pelvic morphological features, the total percentages for each collection of correctly-identified females were quite high. The ventral arc correctly identified 88.2% of Grant females (incorrect for only two females), 90.9% of Lisbon females, 91.2% of Coimbra females, 92.5% of Spitalfields females, 94.4% of Dart females, and 97.2% of Pretoria females. Full details are in Table A2.32 of Appendix 2.

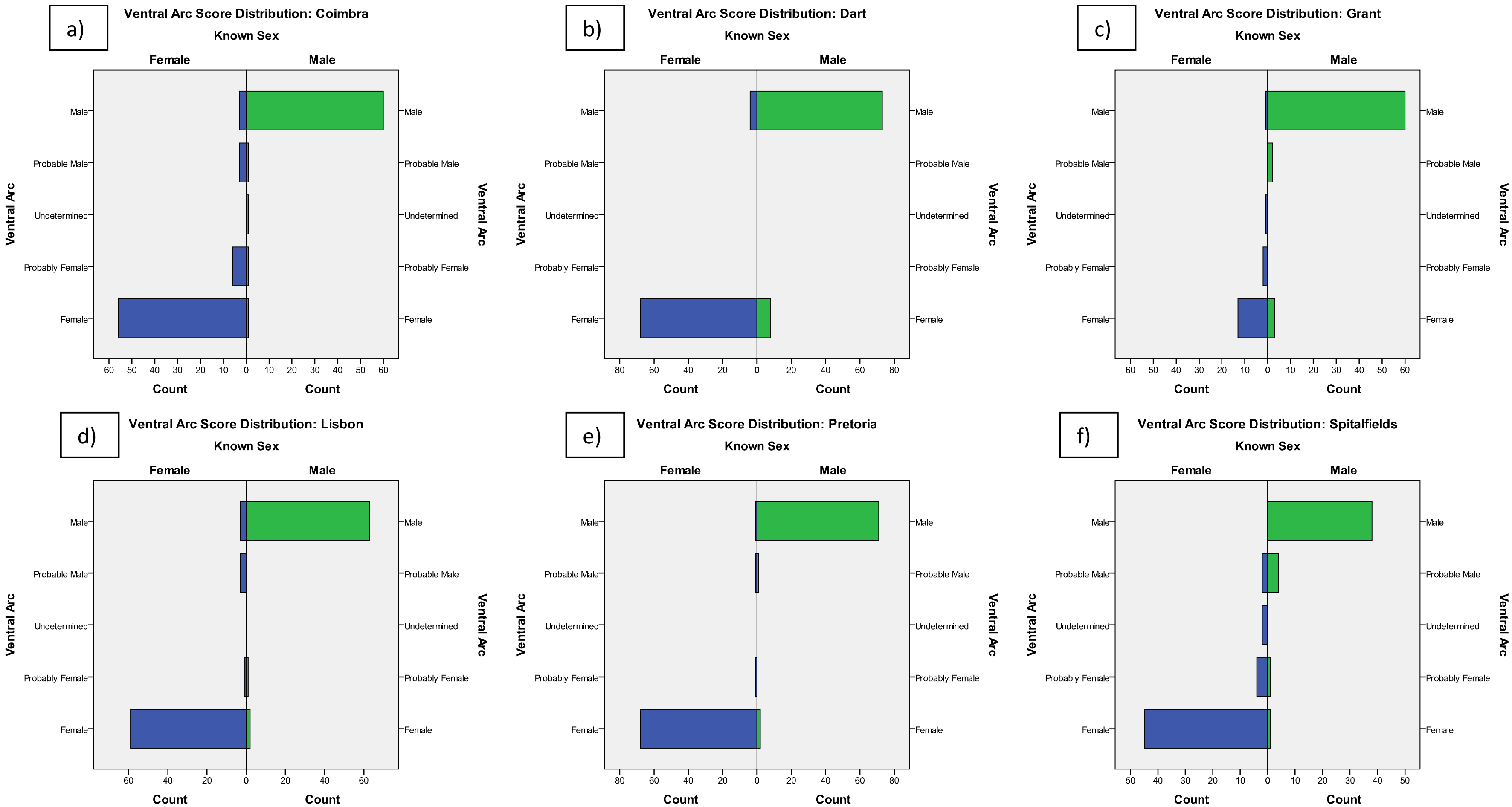
The score distribution as percentages of the sample for each collection (presented in Table 4.30, and see Figures 4.9a to 4.9f for bar charts) were observed to determine the reasons for the significant differences found using the K-S and MWU tests. For females, the only significant differences were in central tendency: for Dart compared to both Coimbra and Grant, and for Pretoria compared to Coimbra, Grant, and Spitalfields. The South African collections had the highest proportions of scores of 1. Meanwhile, Grant, Spitalfields, and Coimbra all had higher proportions of females with scores of 2, 3 (with the exception of Coimbra), 4 (with the exception of Grant), and 5. As such, the South African collections had significantly lower central tendencies than Grant or Coimbra, and Pretoria also had a significantly lower central tendency compared to Spitalfields. In general, scores of 1 were by far the most common. Scores of 2 were the next most common for Grant, Spitalfields and Coimbra. For Lisbon, the second most common scores were 4 and 5. The Pretoria females who did not score 1 were equally divided between scores of 2, 4 and 5. For Dart females, scores of 5 were the second most common. The score distributions of Dart and Pretoria were more highly peaked around scores of 1, while the females from other collections had slightly wider distributions of scores across scoring possibilities.

Score	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	13	76.5	45	84.9	56	82.4	59	89.4	68	94.4	68	95.8
2	2	11.8	4	7.5	6	8.8	1	1.5	0	0.0	1	1.4
3	1	5.9	2	3.8	0	0.0	0	0.0	0	0.0	0	0.0
4	0	0.0	2	3.8	3	4.4	3	4.5	0	0.0	1	1.4
5	1	5.88	0	0.0	3	4.4	3	4.5	4	5.6	1	1.4
Total	17	100.0	53	100.0	68	100.0	66	100.0	72	100.0	71	100.0

Table 4.30. Number of females with each ventral arc score, by collection

For males, the ventral arc again showed no particular age-related trends in the distribution of scores. When low, morphologically “female” scores were present, they were equally likely to occur at younger, middle, or older age groups. There was also no age-related pattern to the percentage of correctly-identified males by age group; the ventral arc was quite good at distinguishing males as well as females, and the lower percentages of correctly-sexed males did not occur in any particular age range. In terms of the total percentage of males sexed correctly for each collection, the values for Grant, Spitalfields, Coimbra, and Lisbon were virtually identical: 95.4%, 95.5%, 95.3% and 95.5%, respectively. Overall, 90.1% of Dart males and 97.3% of Pretoria males were sexed correctly. Full details of proportions of correctly-identified males by age group are in Table A2.33, Appendix 2.

The K-S and MWU tests showed no significant differences in males from any collection, but it is still informative to look at the proportions of scores for each collection (Table 4.31 and Figures 4.9a to 4.9f, below). For all collections, scores of 5 were the most common. Scores of 1 were the next most common for Grant, Lisbon, Dart and Pretoria males, while for Spitalfields, scores of 4 were the next most common. Scores of 2 and 3 were generally uncommon throughout the collections. The scores for Spitalfields were distributed somewhat more widely than for the other collections, which were more highly peaked at scores of 5. Dart’s male ventral arc score distribution was bimodal, as only scores of 1 (9.9%) and 5 (90.1%) were present in the sample.



Figures 4.9a to 4.9f. Ventral arc score distribution bar charts for each collection, separated by sex

Score	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	3	4.6	1	2.3	1	1.6	2	3.0	8	9.9	2	2.7
2	0	0.0	1	2.3	1	1.6	1	1.5	0	0.0	0	0.0
3	0	0.0	0	0.0	1	1.6	0	0.0	0	0.0	0	0.0
4	2	3.1	4	9.1	1	1.6	0	0.0	0	0.0	1	1.4
5	60	92.3	38	86.4	60	93.8	63	95.5	73	90.1	71	95.9
Total	65	100.0	44	100.0	64	100.0	66	100.0	81	100.0	74	100.0

Table 4.31. Number of males with each ventral arc score, by collection

4.4 Overall Sex Determination Results

In general, the morphological methods for sex determination performed well for determining sex for all collections studied. Results were collated by pelvic morphology alone (Phenice, 1969), skull morphology alone (Walker, 2008), and the combined results for the pelvis and skull. The pelvis alone and pelvis and skull together were more successful at determining sex than the skull alone. Results were divided by age and sex group to look for any age- or sex-related trends; finally, sexes and ages were pooled to look at the differences in percentages of individuals correctly identified by sex and by collection. The results of each will be discussed in turn. Tables A3.1 to A3.3 in Appendix 3 provide tables of allocation accuracies (that is, whether sex was estimated correctly) for the pelvis, skull, and pelvis and skull combined by sex and for the sexes pooled.

4.4.1 Pelvis

The pelvis performed well in determining sex for all collections, age groups, and both sexes. When the percentages of correct sex identification are examined by age group, there were no age-related trends – incorrect sex estimation occurred in few cases, and these can be found in any age group. Indeed, the majority of age-sex categories had 100% correct sex identification.

In terms of sex differences and the percentage of correctly-identified individuals, no universal trends were observed. For the Grant Collection, the Lisbon Collection and the Dart Collection, males were more often sexed correctly than females. Grant females were sexed correctly in 94.4% of cases, compared to 96.9% for males; Lisbon females were sexed correctly in 93.4% of cases, compared to 97.1% for males; Dart females were sexed correctly in 91.8% of cases, compared to 97.6% for males. Spitalfields females were sexed correctly more often than were Spitalfields males – 98.5% for females compared to 92.9% for males. For Coimbra and Pretoria, the proportions of correctly-sexed individuals were very similar – 97.3% of Coimbra females were correctly identified compared to 97.0% of Coimbra males, while 97.2% of Pretoria females were correctly identified compared to 97.3% of Pretoria males. Tables 4.32 and 4.33 show the proportions of correctly-sexed females and males, respectively, by age group and collection. For females overall, pelvic morphology performed best for Spitalfields, with 98.5% of females successfully identified, and worst for Dart, with 91.8% of females successfully identified. For males overall, pelvis morphology performed best for Dart, at 97.6%, and worst for Spitalfields, at 92.9%. The values for Lisbon were not far from those of Dart – 93.4% for females compared to 97.1% for males.

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	2	100.0%	8	100.0%	10	100.0%	10	90.0%	10	90.0%	9	100.0%
30-39	1	100.0%	8	100.0%	10	100.0%	9	100.0%	9	77.8%	10	100.0%
40-49	2	100.0%	10	100.0%	10	90.0%	10	90.0%	9	100.0%	10	100.0%
50-59			10	100.0%	10	90.0%	10	90.0%	10	90.0%	10	100.0%
60-69	5	100.0%	10	90.0%	10	100.0%	10	80.0%	8	100.0%	11	90.9%
70-79	5	100.0%	10	100.0%	10	100.0%	10	100.0%	10	80.0%	10	100.0%
80-89	1	100.0%	10	100.0%	10	100.0%	10	100.0%	10	100.0%	10	90.0%
90-99	2	50.0%			3	100.0%	7	100.0%	6	100.0%	2	100.0%
100+					0		0		1	100.0%	0	
Total	18	94.4%	66	98.5%	73	97.3%	76	93.4%	73	91.8%	72	97.2%

Table 4.32. Pelvis sex assessment, females by age and collection
n = number of individuals; % Correct = percentage of correctly-sexed individuals

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	3	100.0%	6	100.0%	9	100.0%	10	100.0%	10	100.0%	10	100.0%
30-39	10	100.0%	9	88.9%	10	90.0%	10	100.0%	10	100.0%	10	90.0%
40-49	9	100.0%	10	90.0%	10	90.0%	10	100.0%	10	90.0%	10	100.0%
50-59	11	90.9%	9	100.0%	10	100.0%	10	100.0%	10	100.0%	10	100.0%
60-69	10	90.0%	8	100.0%	11	100.0%	10	90.0%	10	100.0%	10	100.0%
70-79	13	100.0%	10	90.0%	10	100.0%	10	100.0%	10	100.0%	10	100.0%
80-89	9	100.0%	3	66.7%	6	100.0%	10	90.0%	10	100.0%	10	90.0%
90-99			1	100.0%	1	100.0%	0		9	88.9%	4	100.0%
100+					0		0		3	100.0%	0	
Total	65	96.9%	56	92.9	67	97.0%	70	97.1%	82	97.6%	74	97.3%

Table 4.33. Pelvis sex assessment, males by age and collection
n = number of individuals; % Correct = percentage of correctly-sexed individuals

When the sexes are pooled, no age-related trends became apparent (see Table 4.34, below). Fairly low numbers of incorrectly sexed individuals mean that the proportions of correctly-sexed individuals were fairly high across all age groups – with the exception of the 90 to 99 year olds for the Grant Collection, where only one of the two individuals in this category was sexed correctly (50%). Overall, the pelvis performed well in determining sex for all collections; best for Pretoria, where 97.3% of all individuals were successfully sexed, followed closely by Coimbra, at 97.1%. For both of these collections, the percentages of correctly sexed males and females separately were very close. Next was Spitalfields, at 96.7%, followed by Grant, at 96.4%. Finally, 95.2% of Lisbon individuals were sexed correctly, followed by 94.8% of Dart individuals.

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Corr.	*n*	% Corr.	*n*	% Corr.	*n*	% Corr.	*n*	% Corr.	*n*	% Corr.
20-29	5	100	14	100	19	100	20	95	20	95	19	100
30-39	11	100	17	94	20	95	19	100	19	89	20	95
40-49	11	100	20	95	20	90	20	95	19	95	20	100
50-59	11*	91	19	100	20	95	20	95	20	95	20	100
60-69	15	93	18	94	21	100	20	85	18	100	21	95
70-79	18	100	20	95	20	100	20	100	20	90	20	100
80-89	10	100	13	92	16	100	20	95	20	100	20	90
90-99	2**	50	1*	100	4	100	7**	100	15	93	6	100
100+	0		0		0		0		4	100	0	
Total	83	96.4	122	95.9	140	97.1	146	95.2	155	94.8	146	97.3

Table 4.34. Pelvis sex assessment, sexes pooled, by age and collection
n = number of individuals; % Corr. = percentage of correctly-sexed individuals; * = males only; ** = females only.

4.4.2 Skull

In general, the skull did not perform as well as the pelvis in determining sex. For Grant, Coimbra, Lisbon and Spitalfields, the skull performed fairly well, but was quite poor at determining sex for Dart and Pretoria. As with the pelvis, no age-related trends could be seen – any age group may have a low or high proportion of correctly-sexed individuals.

Unlike the pelvis, there does seem to be a sex-related difference – that is, the skull was better at determining the sex of females than males. For Spitalfields, Coimbra, Lisbon, Dart and Pretoria, the skull better determined sex of females than males; for Grant, the skull performed slightly better for males, but the difference here was minimal (92.3% for females compared to 93.4% for males). For Coimbra, the difference between percentages of correctly-sexed males and females was less than 10%; 89.0% for females compared to 80.6% for males. Echoing these values fairly closely were those of Lisbon, where 92.1% of females were correctly sexed, compared to 81.4% of males. The disparity was larger in the other collections, however; Spitalfields females were correctly sexed in 89.4% of cases, compared to 73.8% of males. Pretoria females were sexed correctly in 70.8% of cases, but for only 49.3% of males. Dart had the largest sex difference; while

92.0% of females were successfully identified, only 48.8% of males were correctly identified. These values represented the lowest male percentage and one of the highest female percentages (Grant had the highest percentage of correct female determinations, at 92.3%). Overall, the skull performed best for Grant Collection females, at 92.3%, followed closely by Lisbon and Dart (92.1% and 92.0%, respectively). Percentages of correct sex determination for Spitalfields and Coimbra were not much less, at 89.4% and 89.0%, respectively; the skull was least successful for determining the sex of Pretoria females, with correct sexing of 70.8%. For males, the skull was most successful again for Grant, at 93.4%; next were Lisbon and Coimbra, at 81.4% and 80.6%, respectively. For Spitalfields, the skull accurately determined sex for 73.8% of males. For Pretoria and Dart males, the skull was not particularly successful at determining sex, at 49.3% and 48.8%, respectively. Tables 4.35 and 4.36 show full details, below.

No particular age-related trends were revealed by pooling the sexes (Table 4.37, below). Unsurprisingly, as Grant had the highest male and female percentages of correct sex determination, the skull was most successful in determining sex for Grant Collection individuals. The skull was fairly successful for Lisbon (87.0%) and Coimbra (85.0%), but slightly less so for Spitalfields (81.7%). It was least successful for Dart and Pretoria individuals, correctly sexing 69.4% and 60.0% of individuals, respectively. It is interesting that the total percentages of correct sex identification were close for the collections located in the same countries – Lisbon and Coimbra, both in Portugal, and Dart and Pretoria, both in South Africa.

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	1	100.0%	9	88.9%	10	100.0%	10	100.0%	10	100.0%	9	88.9%
30-39	1	100.0%	10	80.0%	10	100.0%	9	100.0%	10	90.0%	11	90.9%
40-49	1	100.0%	10	70.0%	10	90.0%	10	90.0%	9	100.0%	10	80.0%
50-59			8	87.5%	10	90.0%	10	90.0%	9	88.9%	11	45.5%
60-69	4	75.0%	10	100.0%	10	80.0%	10	80.0%	10	100.0%	10	90.0%
70-79	3	100.0%	9	100.0%	10	100.0%	10	90.0%	10	70.0%	9	66.7%
80-89	1	100.0%	10	100.0%	10	70.0%	10	90.0%	10	90.0%	10	30.0%
90-99	2	100.0%			3	66.7%	7	100.0%	6	100.0%	2	100.0%
100+					0		0		1	100.0%	0	
Total	13	92.3%	66	89.4%	73	89.0%	76	92.1%	75	92.0%	72	70.8%

Table 4.35. Skull sex assessment, females by age and collection
n = number of individuals; % Correct = percentage of correctly-sexed individuals

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	2	100.0%	6	83.3%	9	55.6%	10	80.0%	10	30.0%	10	50.0%
30-39	9	100.0%	11	54.5%	10	80.0%	10	80.0%	10	50.0%	10	40.0%
40-49	9	100.0%	10	80.0%	10	90.0%	10	80.0%	10	60.0%	10	50.0%
50-59	11	81.8%	10	100.0%	10	80.0%	10	80.0%	10	80.0%	10	60.0%
60-69	9	88.9%	11	63.6%	11	81.8%	10	80.0%	10	30.0%	10	30.0%
70-79	12	91.7%	11	72.7%	10	90.0%	10	80.0%	10	40.0%	9	55.6%
80-89	9	100.0%	5	60.0%	6	83.3%	10	90.0%	10	50.0%	10	60.0%
90-99			1	100.0%	1	100.0%	0		9	44.4%	4	50.0%
100+					0		0		3	66.7%	0	
Total	61	93.4%	65	73.8%	67	80.6%	70	81.4%	82	48.8%	73	49.3%

Table 4.36. Skull sex assessment, males by age and collection
n = number of individuals; % Correct = percentage of correctly-sexed individuals

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Corr.	*n*	% Corr.	*n*	% Corr.	*n*	% Corr.	*n*	% Corr.	*n*	% Corr.
20-29	3	100%	15	87%	19	79%	20	90%	20	65%	19	68%
30-39	10	100%	21	67%	20	90%	19	89%	20	70%	21	67%
40-49	10	100%	20	75%	20	90%	20	85%	19	79%	20	65%
50-59	11*	82%	18	94%	20	85%	20	85%	19	84%	21	52%
60-69	13	85%	21	81%	21	81%	20	80%	20	65%	20	60%
70-79	15	93%	20	85%	20	95%	20	85%	20	55%	18	61%
80-89	10	100%	15	87%	16	75%	20	90%	20	70%	20	45%
90-99	2**	100%	1*	100%	4	75%	7**	100%	15	67%	6	67%
100+	0		0		0		0		4	75%	0	
Total	74	93.2%	131	81.7%	140	85.0%	146	87.0%	157	69.4%	145	60.0%

Table 4.37. Skull sex assessment, sexes pooled, by age and collection
n = number of individuals; % Corr. = percentage of correctly-sexed individuals; * = males only; ** = females only.

4.4.3 Pelvis and Skull Combined

The performance of the pelvis and skull together was generally good, increasing percentages of correctly determined sex over either the skull or pelvis alone (but particularly the skull alone). No age-related trends were observed. In terms of absolute numbers, where individuals were incorrectly sexed in an age group, it was only one, or at most, two individuals (Tables 4.38 and 4.39 provide details). Only a few categories had two incorrectly-sexed individuals; these were 60 to 69 year old Lisbon females, 70 to 79 year old Dart females, and 80 to 89 year old Spitalfields males. In the majority of age groups, 100.0% of individuals were sexed correctly.

No trends by sex were clear, either; for half of the collections (Spitalfields, Coimbra, and Pretoria), females were more often sexed correctly than were males, while for the other half of the collections (Grant, Lisbon, and Dart), males were sexed correctly more often than were females. For females only, the total percentages of correct sex determination ranged from 92.3%, for Grant females, to 98.6% for both Pretoria and Coimbra females. Lisbon females were sexed correctly in 93.4% of cases, Dart females in 94.4% of cases, and Spitalfields females in 98.4% of cases. For males only, the total percentages ranged from 92.9%, for Spitalfields, to 97.6% for Dart. The other percentages were fairly close together; 96.7% of Grant males were correctly identified, compared to 97.0% of Coimbra males, 97.1% of Lisbon males, and 97.3% of Pretoria males.

No age-related trends were observed when the sexes were pooled. Using the pelvis and skull together resulted in a range of proportions of correct sex identifications from 95.2% for Lisbon to 97.9% for both Coimbra and Pretoria individuals. This range was quite narrow; the total percentages of correct sex determination for the other collections were 95.8% for Spitalfields, 96.0% for Grant, and 96.1% for Dart. Table 4.40, below, contains all details.

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	1	100.0%	8	100.0%	10	100.0%	10	90.0%	10	100.0%	9	100.0%
30-39	1	100.0%	8	100.0%	10	100.0%	9	100.0%	9	88.9%	10	100.0%
40-49	1	100.0%	10	100.0%	10	100.0%	10	90.0%	8	100.0%	10	100.0%
50-59		100.0%	8	100.0%	10	90.0%	10	90.0%	9	88.9%	10	100.0%
60-69	4	100.0%	10	90.0%	10	100.0%	10	80.0%	8	100.0%	10	100.0%
70-79	3	100.0%	9	100.0%	10	100.0%	10	100.0%	10	80.0%	9	100.0%
80-89	1	50.0%	10	100.0%	10	100.0%	10	100.0%	10	100.0%	10	90.0%
90-99	2				3	100.0%	7	100.0%	6	100.0%	2	100.0%
100+					0		0		1	100.0%	0	
Total	13	92.3%	63	98.4%	73	98.6%	76	93.4%	71	94.4%	70	98.6%

Table 4.38. Pelvis and skull combined sex assessment, females by age and collection
n = number of individuals; % Correct = percentage of correctly-sexed individuals

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	2	100.0%	6	100.0%	9	100.0%	10	100.0%	10	100.0%	10	100.0%
30-39	9	100.0%	9	88.9%	10	90.0%	10	100.0%	10	100.0%	10	90.0%
40-49	9	100.0%	10	90.0%	10	90.0%	10	100.0%	10	90.0%	10	100.0%
50-59	11	90.9%	9	100.0%	10	100.0%	10	100.0%	10	100.0%	10	100.0%
60-69	9	88.9%	8	100.0%	11	100.0%	10	90.0%	10	100.0%	10	100.0%
70-79	12	100.0%	10	90.0%	10	100.0%	10	100.0%	10	100.0%	9	100.0%
80-89	9	100.0%	3	66.7%	6	100.0%	10	90.0%	10	100.0%	10	90.0%
90-99			1	100.0%	1	100.0%	0		9	88.9%	4	100.0%
100+					0		0		3	100.0%	0	
Total	61	96.7%	57	92.9%	67	97.0%	70	97.1%	82	97.6%	73	97.3%

Table 4.39. Pelvis and skull combined sex assessment, males by age and collection
n = number of individuals; % Correct = percentage of correctly-sexed individuals

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Corr.	*n*	% Corr.	*n*	% Corr.	*n*	% Corr.	*n*	% Corr.	*n*	% Corr.
20-29	3	100.0	14	100.0	19	100.0	20	95.0	20	100.0	19	100.0
30-39	10	100.0	17	94.1	20	95.0	19	100.0	19	94.7	20	95.0
40-49	10	100.0	20	95.0	20	95.0	20	95.0	18	94.4	20	100.0
50-59	11*	90.9	17	100.0	20	95.0	20	95.0	19	94.7	20	100.0
60-69	13	92.3	18	94.4	21	100.0	20	85.0	18	100.0	20	100.0
70-79	15	100.0	19	94.7	20	100.0	20	100.0	20	90.0	18	100.0
80-89	10	100.0	13	92.3	16	100.0	20	95.0	20	100.0	20	90.0
90-99	2**	50.0	1*	100.0	4	100.0	7**	100.0	15	93.3	6	100.0
100+	0		0		0		0		4	100.0	0	
Total	74	96.0	120	95.8	140	97.9	146	95.2	153	96.1	143	97.9

Table 4.40. Pelvis and skull combined sex assessment, sexes pooled, by age and collection
n = number of individuals; % Corr. = percentage of correctly-sexed individuals; * = males only; ** = females only.

In terms of advantages over the use of single skeletal elements for determining sex, the pelvis and skull combined represented an improvement over the skull alone (Tables 4.41, 4.42 and 4.43, below), for females and males separately and for the sexes pooled. The exception was the Grant females, for whom the percentage of correct sex determination was the same using the skull alone or pelvis and skull combined. However, results varied when the success of the pelvis and skull combined was compared to that of the pelvis alone. For Coimbra, Dart and Pretoria, for females alone and for the sexes pooled, the pelvis and skull together gave improved results (percentage of correct sex determination) compared to the pelvis alone. For Coimbra, Spitalfields, Dart and Pretoria, for males alone, the percentages of correct sex determinations were the same using pelvis alone and using pelvis and skull together. For females alone, males alone and sexes pooled, and for Spitalfields females alone and sexes pooled, using the pelvis and skull together actually resulted in slightly lower percentages of correct sex determinations compared to the pelvis alone. However, this decrease was in the order of tenths of a percent. For Lisbon (females and males alone, and the sexes pooled) the pelvis only and pelvis and skull together resulted in exactly the same percentage of correct sex determinations.

Skeletal Element	Grant % Correct	Spitalfields % Correct	Coimbra % Correct	Lisbon % Correct	Dart % Correct	Pretoria % Correct
P	94.4%	98.5%	97.3%	93.4%	91.8%	97.2%
S	92.3%	89.4%	89.0%	92.1%	92.0%	70.8%
P+S	92.3%	98.4%	98.6%	93.4%	94.4%	98.6%

Table 4.41. Females only, total percentages of correct sex identification
P = pelvis only; S = skull only; P+S = pelvis and skull combined.

Skeletal Element	Grant % Correct	Spitalfields % Correct	Coimbra % Correct	Lisbon % Correct	Dart % Correct	Pretoria % Correct
P	96.9%	92.9%	97.0%	97.1%	97.6%	97.3%
S	93.4%	73.8%	80.6%	81.4%	48.8%	49.3%
P+S	96.7%	92.9%	97.0%	97.1%	97.6%	97.3%

Table 4.42. Males only, total percentages of correct sex identification
P = pelvis only; S = skull only; P+S = pelvis and skull combined.

Skeletal Element	Grant % Correct	Spitalfields % Correct	Coimbra % Correct	Lisbon % Correct	Dart % Correct	Pretoria % Correct
P	96.4%	95.9%	97.1%	95.2%	94.8%	97.3%
S	93.2%	81.7%	85.0%	87.0%	69.4%	60.0%
P+S	96.0%	95.8%	97.9%	95.2%	96.1%	97.9%

Table 4.43. Sexes pooled, total percentages of correct sex identification
P = pelvis only; S = skull only; P+S = pelvis and skull combined.

4.5 Albanese's Metrical Sex Method

All modifications were first pooled to examine the overall accuracy of Albanese's (2003a) method for each collection by age group, and in total. Results were then divided by modification number. It was possible to assess sex for the majority of individuals using modification 1, so the accuracy of this modification was examined alone; the results for all other modifications used were pooled because, in most cases, insufficient individuals were sexed using any one modification for results to be meaningfully compared. Modification 1 performed better than the other modifications combined in most cases.

For all modifications taken together, no age-related trends could be seen; that is, the metrical method worked equally well on all age groups, and errors in sex determination occurred in any age group. When males and females were separated (see Tables 4.44 and 4.45), and absolute numbers of errors were examined, only one or two misclassified individuals tended to occur in any particular age group. The highest number of incorrect sex determinations was four, occurring in the 70 to 79 age group of Pretoria males, resulting in only 60% of the group being correctly sexed.

Overall, Albanese's metrical method performed better for Spitalfields, Coimbra, and Lisbon females compared to males. However, for Dart and Pretoria, more males were correctly identified than females. For the Grant Collection, all females and all males were sexed correctly. The metrical method performed best for the Grant Collection (100% of individuals were correctly sexed), followed by Spitalfields (94%) (Table 4.46 has details). It performed moderately well for the other collections – 92.9% of Coimbra individuals were correctly sexed, as were 92.1% of both Lisbon and Dart individuals, and 90.2% of Pretoria individuals. While these percentages seem reasonable, that there was quite a disparity between sex determination for males and females from the Dart, Lisbon, and particularly, the Coimbra collections is more problematic. For Dart and Lisbon, nearly 10% more females were sexed correctly compared to males; for Coimbra, the difference is 15.5% (in favour of females).

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	2	100.0%	5	100.0%	9	100.0%	10	100.0%	9	88.9%	9	100.0%
30-39	1	100.0%	7	100.0%	10	100.0%	9	100.0%	8	87.5%	10	90.0%
40-49	2	100.0%	7	100.0%	9	100.0%	9	88.9%	9	100.0%	10	90.0%
50-59	0		5	100.0%	10	100.0%	9	88.9%	10	90.0%	10	90.0%
60-69	5	100.0%	8	100.0%	8	100.0%	7	85.7%	8	87.5%	10	100.0%
70-79	5	100.0%	4	75.0%	9	100.0%	8	100.0%	10	80.0%	8	75.0%
80-89	1	100.0%	8	100.0%	10	100.0%	9	100.0%	10	100.0%	10	80.0%
90-99	2	100.0%	0		3	100.0%	5	100.0%	6	66.7%	2	100.0%
100+	0		0		0		0		1	0.0%	0	
Total	18	100.0%	44	97.7%	68	100.0%	66	95.5%	71	87.3%	69	89.9%

Table 4.44. Percentages of correctly sexed females, all modifications together

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	3	100.0%	4	100.0%	6	100.0%	9	88.9%	9	100.0%	10	100.0%
30-39	10	100.0%	7	71.4%	7	71.4%	9	100.0%	10	100.0%	10	90.0%
40-49	9	100.0%	6	100.0%	10	80.0%	10	100.0%	10	90.0%	10	100.0%
50-59	11	100.0%	7	100.0%	10	90.0%	10	90.0%	9	100.0%	10	90.0%
60-69	10	100.0%	5	80.0%	10	80.0%	8	75.0%	10	90.0%	10	100.0%
70-79	13	100.0%	4	100.0%	8	87.5%	9	77.8%	10	100.0%	10	60.0%
80-89	9	100.0%	1	100.0%	6	83.3%	6	83.3%	10	100.0%	10	90.0%
90-99	0		1	100.0%	1	100.0%	0		9	88.9%	4	100.0%
100+	0		0		0		0		3	100.0%	0	
Total	65	100.0%	35	91.4%	58	84.5%	61	88.5%	80	96.3%	74	90.5%

Table 4.45. Percentages of correctly sexed males, all modifications together

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	5	100.0%	9	100.0%	15	100.0%	19	94.7%	18	94.4%	19	100.0%
30-39	11	100.0%	14	85.7%	17	88.2%	18	100.0%	18	94.4%	20	90.0%
40-49	11	100.0%	13	100.0%	19	89.5%	19	94.7%	19	94.7%	20	95.0%
50-59	11	100.0%	12	100.0%	20	95.0%	19	89.5%	19	94.7%	20	90.0%
60-69	15	100.0%	13	92.3%	18	88.9%	15	80.0%	18	88.9%	20	100.0%
70-79	18	100.0%	8	87.5%	17	94.1%	17	88.2%	20	90.0%	18	66.7%
80-89	10	100.0%	9	100.0%	16	93.8%	15	93.3%	20	100.0%	20	85.0%
90-99	2	100.0%	1	100.0%	4	100.0%	5	100.0%	15	80.0%	6	100.0%
100+	0		0		0		0		4	75.0%	0	
Total	83	100.0%	79	94.9%	126	92.9%	127	92.1%	151	92.1%	143	90.2%

Table 4.46. Percentages of correctly sexed individuals, all modifications together, sexes pooled

Modification 1 was next analysed alone, as this was Albanese's best-fit model, and used the most measurements (Table 4.47 for females only, Table 4.48 for males only, and Table 4.49 for the sexes pooled). Again, there were no age-related trends in ability to predict sex using this method. In absolute terms, few individuals were sexed incorrectly in any one age category (generally only one or two incorrect sex predictions per group where errors occurred). Again, females were more often sexed correctly for Spitalfields, Coimbra and Lisbon compared to males, and vice versa for Dart and Pretoria. While Albanese's method had the same success rate for sexing Grant males and females, and very similar rates for Spitalfields and Pretoria males and females, the disparity between male and female results using modification 1 for the other collections is again somewhat problematic. For Coimbra, the difference was 15.9%, and for both Lisbon and Dart, it was 11.1%. In terms of overall sex determination rates, modification 1 was most successful for Grant, for which 100.0% of individuals were correctly sexed, followed by Spitalfields, for which 95.9% of individuals were correctly identified. For Lisbon, 94.6% of individuals were correctly sexed, as were 93.5% of Dart individuals, 92.9% of Coimbra individuals and 89.6% of Pretoria individuals.

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	1	100.0%	1	100.0%	7	100.0%	6	100.0%	9	88.9%	9	100.0%
30-39	1	100.0%	5	100.0%	9	100.0%	6	100.0%	8	87.5%	9	88.9%
40-49	2	100.0%	5	100.0%	8	100.0%	7	100.0%	8	100.0%	9	88.9%
50-59	0		3	100.0%	9	100.0%	6	100.0%	10	90.0%	10	90.0%
60-69	5	100.0%	6	100.0%	6	100.0%	3	100.0%	7	85.7%	10	100.0%
70-79	5	100.0%	2	50.0%	8	100.0%	4	100.0%	9	88.9%	7	71.4%
80-89	1	100.0%	5	100.0%	6	100.0%	3	100.0%	7	100.0%	9	77.8%
90-99	0		0		2	100.0%	3	100.0%	5	60.0%	1	100.0%
100+	0		0		0		0		1	0.0%	0	
Total	15	100.0%	27	96.3%	55	100.0%	38	100.0%	64	87.5%	64	89.1%

Table 4.47. Percentages of correctly sexed females, modification 1

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	3	100.0%	3	100.0%	6	100.0%	6	83.3%	9	100.0%	10	100.0%
30-39	10	100.0%	5	80.0%	5	60.0%	6	100.0%	9	100.0%	10	90.0%
40-49	7	100.0%	3	100.0%	7	85.7%	6	100.0%	9	100.0%	9	100.0%
50-59	10	100.0%	5	100.0%	8	100.0%	8	87.5%	8	100.0%	10	90.0%
60-69	8	100.0%	3	100.0%	9	77.8%	6	83.3%	10	90.0%	9	100.0%
70-79	11	100.0%	3	100.0%	3	66.7%	1	100.0%	10	100.0%	10	60.0%
80-89	9	100.0%	0		5	80.0%	3	66.7%	10	100.0%	9	88.9%
90-99	0		0		1	100.0%	0		7	100.0%	4	100.0%
100+	0		0		0		0		2	100.0%	0	
Total	58	100.0%	22	95.5%	44	84.1%	36	88.9%	74	98.6%	71	90.1%

Table 4.48. Percentages of correctly sexed males, modification 1

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	4	100.0%	4	100.0%	13	100.0%	12	91.7%	18	94.4%	19	100.0%
30-39	11	100.0%	10	90.0%	14	85.7%	12	100.0%	17	94.1%	19	89.5%
40-49	9	100.0%	8	100.0%	15	93.3%	13	100.0%	17	100.0%	18	94.4%
50-59	10	100.0%	8	100.0%	17	100.0%	14	92.9%	18	94.4%	20	90.0%
60-69	13	100.0%	9	100.0%	15	86.7%	9	88.9%	17	88.2%	19	100.0%
70-79	16	100.0%	5	80.0%	11	90.9%	5	100.0%	19	94.7%	17	64.7%
80-89	10	100.0%	5	100.0%	11	90.9%	6	83.3%	17	100.0%	18	83.3%
90-99	0		0		3	100.0%	3	100.0%	12	83.3%	5	100.0%
100+	0		0		0		0		3	66.7%	0	
Total	73	100.0%	49	95.9%	99	92.9%	74	94.6%	138	93.5%	135	89.6%

Table 4.49. Percentages of correctly sexed individuals, sexes pooled, modification 1

The success of all the other modifications, numbering 2 to 26, were analysed together, as the absolute numbers of individuals for any one of those modifications alone were low (Tables 4.50 to 4.52). As for modification 1, no age-related trends could be seen, and absolute numbers of incorrectly-sexed individuals for any particular age category were low, with one or two errors per age group, if any. For Spitalfields, Coimbra, Lisbon, and Dart, these collected modifications were more successful for females than males; 100.0% for Spitalfields females, compared to 84.6% for Spitalfields males, 100.0% for Coimbra females compared to 85.7% for Coimbra males, 89.3% for Lisbon females compared to 88.0% for Lisbon males, and 85.7% for Dart females compared to 66.7% for Dart males. For both Grant and Pretoria, 100.0% of both males and females were sexed correctly using modifications 2 to 26; these modifications were most successful at sexing Grant and Pretoria individuals. Modifications 2 to 26 were fairly successful for sex determination of Spitalfields and Coimbra individuals overall, at 93.3% and 92.6%, respectively, followed by Lisbon, at 88.7%. They were least successful for Dart individuals, with correct sex determination in 76.9% of cases.

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	1	100.0%	4	100.0%	2	100.0%	4	100.0%	0		0	
30-39	0		2	100.0%	1	100.0%	3	100.0%	0		1	100.0%
40-49	0		2	100.0%	1	100.0%	2	50.0%	1	100.0%	1	100.0%
50-59	0		2	100.0%	1	100.0%	3	66.7%	0		0	
60-69	0		2	100.0%	2	100.0%	4	75.0%	1	100.0%	0	
70-79	0		2	100.0%	1	100.0%	4	100.0%	1	0.0%	1	100.0%
80-89	0		3	100.0%	4	100.0%	6	100.0%	3	100.0%	1	100.0%
90-99	2	100.0%	0		1	100.0%	2	100.0%	1	100.0%	1	100.0%
100+	0		0		0		0		0		0	
Total	3	100.0%	17	100.0%	13	100.0%	28	89.3%	7	85.7%	5	100.0%

Table 4.50. Percentages of correctly sexed females, all other modifications (2 to 26)

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	0		1	100.0%	0		3	100.0%	0		0	
30-39	0		2	50.0%	2	100.0%	3	100.0%	1	100.0%	0	
40-49	2	100.0%	3	100.0%	3	66.7%	4	100.0%	1	0.0%	1	100.0%
50-59	1	100.0%	2	100.0%	2	50.0%	2	100.0%	1	100.0%	0	
60-69	2	100.0%	2	50.0%	1	100.0%	2	50.0%	0		1	100.0%
70-79	2	100.0%	1	100.0%	5	100.0%	8	75.0%	0		0	
80-89	0		1	100.0%	1	100.0%	3	100.0%	0		1	100.0%
90-99	0		1	100.0%	0		0		2	50.0%	0	
100+	0		0		0		0		1	100.0%	0	
Total	7	100.0%	13	84.6%	14	85.7%	25	88.0%	6	66.7%	3	100.0%

Table 4.51. Percentages of correctly sexed males, all other modifications (2 to 26)

Age Group	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
	n	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct	*n*	% Correct
20-29	1	100.0%	5	100.0%	2	100.0%	7	100.0%	0		0	
30-39	0		4	75.0%	3	100.0%	6	100.0%	1	100.0%	1	100.0%
40-49	2	100.0%	5	100.0%	4	75.0%	6	83.3%	2	50.0%	2	100.0%
50-59	1	100.0%	4	100.0%	3	66.7%	5	80.0%	1	100.0%	0	
60-69	2	100.0%	4	75.0%	3	100.0%	6	66.7%	1	100.0%	1	100.0%
70-79	2	100.0%	3	100.0%	6	100.0%	12	83.3%	1	0.0%	1	100.0%
80-89	0		4	100.0%	5	100.0%	9	100.0%	3	100.0%	2	100.0%
90-99	2	100.0%	1	100.0%	1	100.0%	2	100.0%	3	66.7%	1	100.0%
100+	0		0		0		0		1	100.0%	0	
Total	10	100.0%	30	93.3%	27	92.6%	53	88.7%	13	76.9%	8	100.0%

Table 4.52. Percentages of correctly sexed individuals, sexes pooled, all other modifications (2 to 26)

Modification 1 performed better than the other (2 to 26) modifications for most collections, supporting its use as the best fit model. Indeed, for Spitalfields, Coimbra, Lisbon and Grant, the overall results for modification 1 were better than the pooled overall results for modifications 2 to 26. For Grant, the results were the same, at 100.0%. For Pretoria, the results for the other modifications were better than those for modification 1. For females only, modification 1 had the same results as the other modifications for Grant and Coimbra, but for Spitalfields and Pretoria females, the other modifications were more successful in determining sex than was modification 1. For Lisbon and Dart, modification 1 provided better results for females only than the other modifications combined. For males alone, modification 1 gave the same result as did the other modifications for the Grant Collection. For Spitalfields, Lisbon and Dart, modification 1 provided better results than the other modifications combined. For Coimbra and Pretoria, the other modifications provided better results than modification 1. It is worth noting that for collections where the other modifications performed better than modification 1, sample sizes for the other modifications were small.

Independent samples t-tests were done between males and females of each collection to see whether the measurements taken did indeed display sexual dimorphism (see Table A2.15, Appendix 2 for p-values). Obviously, it is desirable that a measurement will be significantly different between the sexes to allow for sex discrimination; however, the t-test is just measuring the mean, and if the male and female means are not significantly different, it does not necessarily mean that the distributions are not bimodal (that is, cannot discriminate sex). The majority of the measurements showed significant differences in mean between the sexes for all the collections tested; only the iliac breadth and SPRL showed non-significant differences. For iliac breadth, significant differences were found between the sexes for the Coimbra and Grant Collections, with the differences in mean between Spitalfields males and females approaching significance. For SPRL, significant differences in mean were found between males and females of the Dart, Lisbon, Pretoria and Spitalfields collections (but not Coimbra and Grant).

Following Jordana et al. (2010: 678), the percentage of sexual dimorphism was calculated for each measurement for each collection, using the formula (% sex difference) = [(male – female)/female] x 100. This allowed comparison of the proportion of sexual dimorphism for each measurement between collections. Table 4.53 shows the results. For the maximum length, maximum diameter and epicondylar breadth of the femur, hip bone height and AIL, the amount of sexual dimorphism was more similar between Coimbra and Lisbon and between Dart and Pretoria compared to the other collections. For example, for the maximum length of the femur, the amount of sexual dimorphism for Coimbra and Lisbon was 8.86% and 8.59%, respectively, and for Dart and Pretoria, 6.53% and 5.63%, respectively. For Grant, sexual dimorphism was 10.38%, and for Spitalfields, 6.74%. For iliac breadth, Dart and Pretoria still clustered together, but Lisbon and Coimbra did not. Similarly, for SPRL, the amount of sexual dimorphism for Dart and Pretoria were fairly close, but Lisbon and Coimbra values did not cluster. Meanwhile, the amount of sexual dimorphism for Grant and Spitalfields were more independent. Grant had the most sexual dimorphism for all measurements except SPRL, for which the greatest amount of sexual dimorphism was found in the Lisbon Collection. For the maximum length and maximum diameter of the femur, the amount of sexual dimorphism for Spitalfields was intermediate to the clusters of Coimbra/Lisbon and Dart/Pretoria. For epicondylar breadth, Spitalfields' values were between those of Dart and Pretoria, while for hip bone height, Spitalfields had the second highest amount of sexual dimorphism (after that of Grant). For SPRL, Spitalfields displayed intermediate values, while for AIL, the Spitalfields values were closest to those of Coimbra and Lisbon.

	Femur Max Length	Femur Max Diameter	Femur Epicondylar Breadth	Hip Bone Height	Iliac Breadth	SPRL	AIL
Coimbra	8.86%	13.03%	10.70%	7.21%	2.28%	-1.89%	10.14%
Dart	6.53%	11.55%	8.86%	6.26%	1.84%	-3.90%	7.61%
Grant	10.38%	15.92%	13.20%	12.12%	6.20%	-2.29%	13.61%
Lisbon	8.59%	11.36%	11.41%	7.96%	0.95%	-6.25%	9.35%
Pretoria	5.63%	11.69%	9.92%	4.84%	1.53%	-5.05%	6.62%
Spitalfields	6.74%	12.04%	9.08%	9.24%	2.65%	-3.26%	9.44%

Table 4.53. Percentages of sexual dimorphism for each measurement, between the sexes for each collection

4.6 Age Estimation Methods

Two auricular surface age estimation methods (developed by Meindl and colleagues and Buckberry and Chamberlain), one pubic symphysis method (developed by Brooks and Suchey), the sternal end of the fourth rib (developed by İşcan and Loth), and cranial suture closure (the method developed by Meindl and colleagues), were used to estimate age-at-death for the individuals sampled from each collection. The individual methods performed variably for the collections in terms of precision of age estimates and relationship with age; the pelvic methods were more useful in age determination than were cranial suture closure and the fourth rib, and the subjective method gave more accurate (and precise) results than did the overall method. Accuracy refers to whether the estimated age range encompassed the actual age of an individual.

Results for each method on its own were analysed by sex, to look for differences between males and females in each collection, which could perhaps suggest whether separate standards for males and females were necessary. The pubic symphyseal ageing method of Suchey and Brooks, for example, uses separate male and female standards, with different male and female age ranges associated with each phase. Comparisons were also made between collections, to look for variation in ageing rates – one-way ANOVAs were used for this purpose, testing the mean ages by collection for differences by each phase or score. The Kolmogorov-Smirnov test was used to look for differences in the cumulative distributions of phase or score for each collection compared to each other collection; this tested the “location” of ageing differences or the shape of the ageing curve, rather than rate. The results were further subdivided into ten-year age groups to see where any differences between collections were located over the life course. Another data subdivision was possible for the Buckberry-Chamberlain auricular surface method; as this method involves scoring specific characteristics or qualities of the auricular surface, before combining the scores for an ultimate score and age estimate, it was possible to analyse the distributions and rates of ageing for each characteristic alone, to further compare the ageing characteristics between collections. In the sections that follow, the results for each of these will be detailed.

4.7 Sex Differences in Ageing According to Method – Within Collection

The equality of means and variances for each sex, divided by collection and score or phase, were tested in order to look for any sex differences. Some significant differences were found with respect to particular phases for some collections. However, in no instance for any collection were all phases significantly different in mean or variance between the sexes.

For the Suchey-Brooks pubic symphysis method, there were statistically significant differences between males and females in only two collections and in two phases. The mean ages for Phase 2 between males and females from the Dart Collection and from the Coimbra Collection, and in variance for Phase 3 Dart Collection males and females were statistically significant. As no clear trends were found, the detailed results are presented in Tables A4.1 to A4.6 in Appendix 4 for each method and collection.

For the Meindl-Lovejoy auricular surface method, only Pretoria males and females in the 40-44 and 45-49 year phases showed significant differences in mean age.

For the Buckberry-Chamberlain auricular surface method, Coimbra males and females differed significantly in variance for phases 2 and 7 and in mean for phase 3. Pretoria males and females had a significant difference in variance in phase 2, and Dart males and females had a significant difference in variance in phase 7.

For lateral-anterior cranial suture phases, the only significant differences between the sexes were for the Lisbon Collection, in variance and mean for phase 5, and in variance only for phase 6. For vault suture phases, a significant difference in mean between Coimbra males and females was found in phase 3, between Dart males and females for phase 6, and in variance between Lisbon males and females for phase 3.

For the sternal end of the fourth rib, significant differences between Coimbra males and females were found in phase 9 (variance only) and in phase 10 (mean only), in phase 10 between males and females from Lisbon (mean) and Pretoria (variance). Other significant differences were found between Pretoria males and females in phase 6 (variance) and phase 8 (mean).

As the Buckberry-Chamberlain method requires summing scores of five morphological traits into a composite score for the final phase association, the scores for these traits were also subjected to tests of equality of variance and mean by sex for each collection. As with the phases for each method, no clear trends were found in terms of significant differences between the sexes for any particular collection or scored trait. For instance, for transverse organisation, the only significant differences between males and females were found in scores of 5 for Spitalfields and Coimbra. As no relevant patterns emerged, the detailed results are presented in Tables A4.7 to A4.11 in Appendix 4 for each scored trait and collection.

4.8 Variation in Phase/Score Distribution

Each combination of collections was tested for variation in phase or score distribution for each ageing method using the two-sample K-S test. The MWU test was also used to test for differences in the median values of the distributions. Collections were tested with the sexes pooled together, females only and males only. The results are presented below, to three decimal places.

4.8.1 Suchey-Brooks Pubic Symphysis Method

The Suchey-Brooks pubic symphysis method was tested first with the sexes pooled. Differences in median between the Grant and Lisbon and the Grant and Spitalfields collections approached significance. Table 4.54 below has the results of the statistical tests.

For females (Table A5.1, Appendix 5), the only significant differences were in median values between Grant and every other collection; Grant's skewed female age distribution (few younger females) was the reason. However, when males were examined alone (Table 4.55), significant differences in distribution and median were found between Grant and Lisbon, and in median only between Grant and Spitalfields, Lisbon and Dart, and Lisbon and Pretoria.

The Grant Collection differences in median and distribution for the pooled results were largely due to the low numbers of individuals in the first three phases, but particularly in phase III, and higher proportions of individuals in phases V and VI compared to the other collections (see Table 4.56 and Figure 4.10 for phase frequencies, below). For all collections except Spitalfields and Lisbon, the highest frequencies of individuals belonged to phase V, and for Spitalfields and Lisbon, the highest frequency was in phase IV. The female frequency patterning followed that of the pooled results.

In terms of males alone, the significant differences found in median in Grant compared to Lisbon mirrored the differences found in the pooled results, exacerbated by a particularly low proportion of Lisbon males in phase VI. The significant differences in median between Lisbon and the South African collections were due to the higher proportions of males in phases V and VI and the lower proportion of males in phase IV in the Dart and Pretoria collections compared to Lisbon, leading to their higher median values. The frequency patterning for males alone followed that of the pooled results.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.805	.450	.018	.283	.139	1.000	.988	.204	.309
Dart			.287	.006	.394	.160	.999	.807	.285	.340
Grant					.006	**.000**	.251	.015	.005	**.002**
Lisbon							.468	.118	.986	.775
Pretoria									.340	.278

Table 4.54. Suchey-Brooks, sexes pooled: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.989	.841	.688	.089	.083	.107	.869	.557	.691	.500
Dart			.913	.102	.105	.042	.926	.698	.643	.313
Grant					.008	**.001**	.740	.266	.214	.032
Lisbon							.104	.034	.929	.749
Pretoria									.534	.216

Table 4.55. Suchey-Brooks, males only: two-sample K-S and MWU results

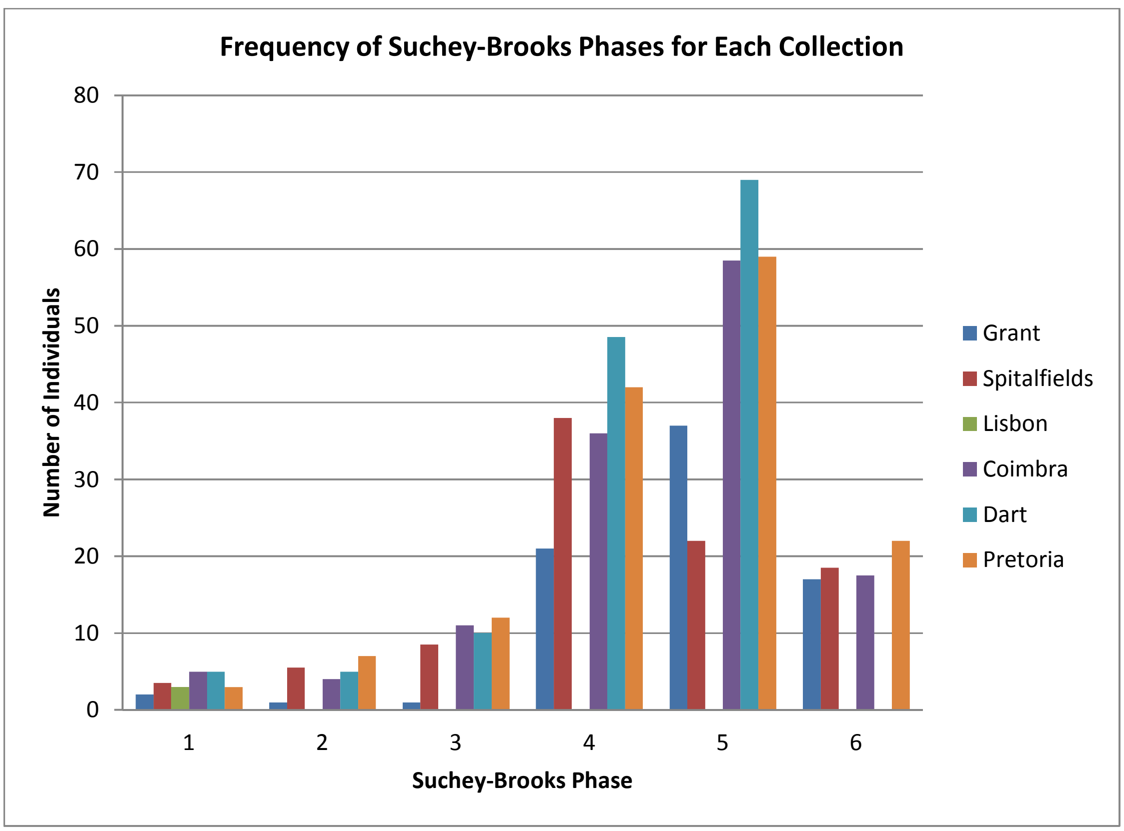


Figure 4.10. Bar chart of Suchey-Brooks phase frequency for each collection, sexes pooled

	Phase																	
	1			**2**			**3**			**4**			**5**			**6**		
	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**
Coimbra	1.5	3.5	**5**	3.5	0.5	**4**	4	7	**11**	21	15	**36**	24	35	**58.5**	14	3.5	**17.5**
Dart	2	3	**5**	3	2	**5**	7	3	**10**	22	27	**48.5**	28	42	**69**	10	5.5	**15.5**
Grant	1	1	**2**	0	1	**1**	0	1	**1**	2	19	**21**	4	33	**37**	7	10	**17**
Lisbon	2	1	**3**	4.5	4	**8.5**	9.5	5	**14.5**	16	31	**47**	20	22	**42**	14	4	**18**
Pretoria	2	1	**3**	5	2	**7**	4	8	**12**	23	20	**42**	27	33	**59**	11	11	**22**
Spitalfields	0.5	3	**3.5**	4.5	1	**5.5**	3	5.5	**8.5**	25	14	**38**	8.5	14	**22**	13	5.5	**18.5**

Table 4.56. Suchey-Brooks phase frequency by collection
F: female; M: male; P: pooled.

4.8.2 Meindl-Lovejoy Auricular Surface Method

The auricular surface methods were tested next, beginning with the Meindl-Lovejoy method. The results for the sexes pooled showed significant differences in score distribution and median between the Grant Collection and every other collection except in distribution compared to Spitalfields. The Spitalfields Collection had a significantly different median compared to every other collection, and a significantly different distribution compared to Dart and Pretoria (and nearing significance compared to the other collections). No other significant differences were found (see Table 4.57, below). These differences remained largely the same for females only and males only (see Tables 4.58 and 4.59).

The phase frequencies (found in Table 4.60 and Figure 4.11, below), revealed the reasons for the significant differences for Grant and Spitalfields in phase distribution and higher median compared to the other collections. Upon examination of the pooled phase frequencies, it was seen that Grant and Spitalfields both had low proportions of individuals in the first four phases compared to the other collections. These differences were magnified by particularly low proportions of Grant individuals in the 25-29 and 30-34 year phases. Both Grant and Spitalfields had relatively high proportions of individuals in the oldest age phases compared to the other collections. While both displayed the same patterning of results, Grant was more extreme, with higher proportions in the oldest age phases and lower numbers in the youngest age phases, resulting in the significant differences between Grant and Spitalfields.

The females-only phase frequencies followed the same patterning as that of the pooled results, reflected in the significant differences found in Grant and Spitalfields females compared to females from the other collections. The same was also true for the males alone, except that no significant differences were found in distribution between Spitalfields and Grant or Spitalfields and Coimbra (distribution differences neared significance between Spitalfields compared to Lisbon and Pretoria); this was because the proportions of Coimbra males in the higher phases (except 60+) were slightly higher than for the other collections. The Grant males-only results were not quite as extreme as those of Grant females, so only the median was significantly different compared to Spitalfields.

In general, the most common Meindl-Lovejoy phase was variable; for Coimbra and Lisbon, the highest number of individuals was in the 30-34 year phase. For Dart and Pretoria, the highest number of individuals was in the 35-39 year phase, while for Spitalfields and Grant, the highest number of individuals was in the 50-60 year phase. The South African collections had low numbers of individuals in the three highest phases. The Pretoria and Dart phase distributions were more highly peaked than those of the other collections; Coimbra, Grant, Lisbon and Spitalfields had flatter, more evenly distributed phase distributions.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.223	.268	**.000**	**.000**	.941	.247	.137	.101	.024	**.000**
Dart			**.000**	**.000**	.431	.946	.965	.518	**.000**	**.000**
Grant					**.000**	**.000**	**.000**	**.000**	.045	**.002**
Lisbon							.383	.688	.007	**.000**
Pretoria									**.000**	**.000**

Table 4.57. Meindl-Lovejoy, sexes pooled: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.593	.324	**.000**	**.000**	.819	.476	.127	.066	.158	.018
Dart			**.000**	**.000**	.482	.628	.994	.356	**.002**	**.001**
Grant					**.000**	**.000**	**.000**	**.000**	.019	**.003**
Lisbon							.119	.170	.043	**.002**
Pretoria									**.000**	**.000**

Table 4.58. Meindl-Lovejoy, females only: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.718	.530	**.001**	**.000**	.926	.291	.760	.618	.199	.011
Dart			**.000**	**.000**	.288	.470	.999	.980	**.005**	**.001**
Grant					**.000**	**.000**	**.000**	**.000**	.323	.034
Lisbon							.683	.426	.014	**.001**
Pretoria									.005	**.001**

Table 4.59. Meindl-Lovejoy, males only: two-sample K-S and MWU results

	Phase																							
	20-24			25-29			30-34			35-39			40-44			45-49			50-60			60+		
	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**
Coimbra	1	1	**2**	15	12	**27**	21	16	**37**	17	18	**35**	13	19	**32**	5	9	**14**	14	7	**21**	4	0	**4**
Dart	3	3	**6**	17	9	**26**	15	14	**29**	29	35	**64**	15	21	**36**	4	2	**6**	5	5	**10**	4	1	**5**
Grant	1	0	**1**	1	0	**1**	0	3	**3**	0	14	**14**	1	12	**13**	3	16	**19**	7	13	**20**	5	4	**9**
Lisbon	6	3	**9**	14	16	**30**	16	18	**34**	13	18	**31**	18	13	**31**	10	6	**16**	9	5	**14**	0	4	**4**
Pretoria	5	2	**7**	13	9	**22**	21	22	**43**	27	33	**60**	15	18	**33**	3	6	**9**	3	8	**11**	1	0	**1**
Spitalfields	1	0	**1**	5	4	**9**	14	9	**23**	13	10	**23**	9	15	**24**	8	10	**18**	20	9	**29**	4	2	**6**

Table 4.60. Meindl-Lovejoy phase frequency by collection

F: female; M: male; P: pooled.

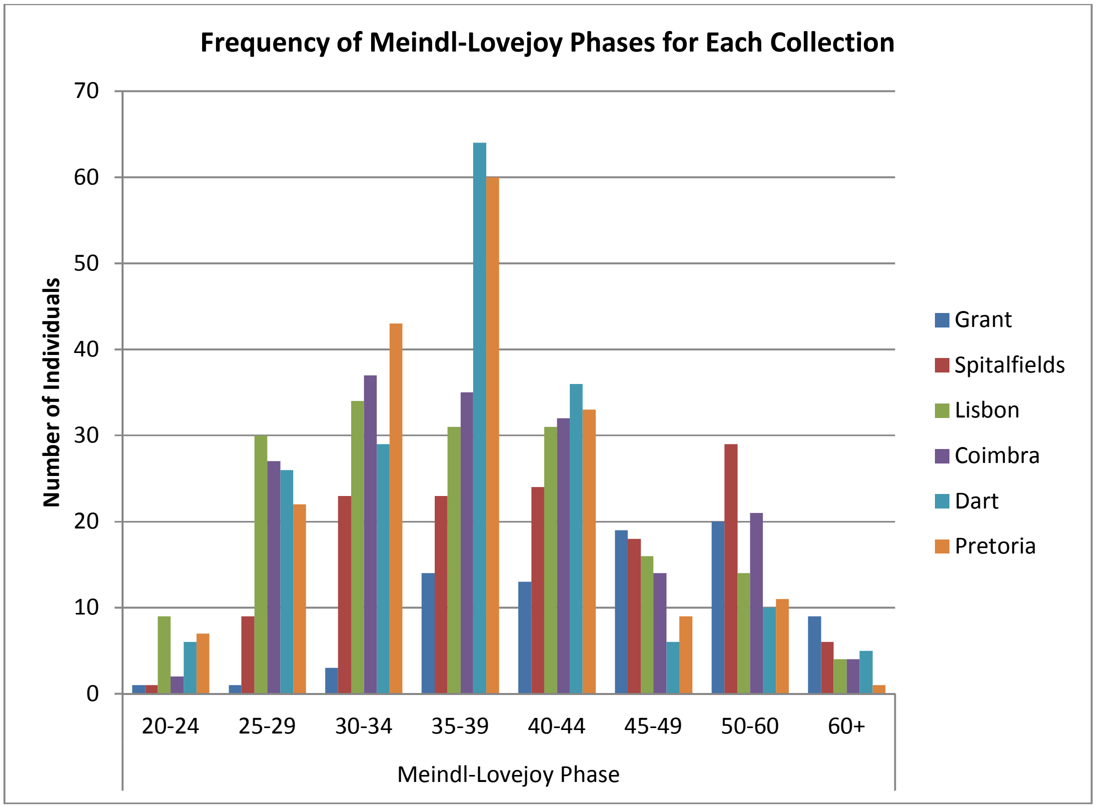


Figure 4.11. Bar chart of Meindl-Lovejoy phase frequency for each collection, sexes pooled

4.8.3 Buckberry-Chamberlain Auricular Surface Method

The Buckberry-Chamberlain auricular surface method pooled results showed similar trends to the Meindl-Lovejoy results (see Tables 4.61 to 4.63, below). Here, too, the Grant and Spitalfields collections showed significant differences in phase distribution and median compared to all other collections (except between Grant and Spitalfields, but this neared significance, and between Spitalfields and Coimbra). Coimbra and Pretoria also had significant differences in phase distribution and median. For females only, Grant was significantly different compared to all other collections (although in median only compared to Spitalfields and Coimbra). Spitalfields females were significantly different in phase and distribution compared to Pretoria females, and in median compared to Dart; compared to Lisbon females, the difference in median neared significance. Coimbra and Pretoria females were significantly different in median and neared significance in distribution. For males only, Grant was significantly different compared to Lisbon and Pretoria in distribution and median, and in median only compared to Dart. Spitalfields was significantly different in median compared to Pretoria. Other differences neared significance.

Phase frequencies for the pooled results showed similar patterns to those found with the Meindl-Lovejoy results (see Table 4.64 and Figure 4.12). Grant and Spitafields again had absolutely and relatively low numbers of individuals in the lower age phases and higher proportions of individuals in the higher phases; Grant's distribution was again more extreme, particularly in the oldest age phase. Significant differences between Pretoria and Coimbra were due to Pretoria's high proportion of individuals in phases III and IV, and low proportions of individuals in phase VII compared to Coimbra.

The differences between Grant Collection females compared to females from the other collections were for the same reasons as for the sexes pooled, again exacerbated by the fact that no Grant females were in the lowest phases. The small sample size for Grant females was not helpful, but the frequencies followed the same patterning as that for Grant males. As before, the significant differences between Spitalfields and the other collections followed the same pattern as that for Grant, but were less extreme, resulting in the significantly different medians. Pretoria's high proportions of females in phases III and IV and lower proportions of females in the higher phases contributed to the significant differences compared to Coimbra and Spitalfields. The significant differences found between collections for males only occurred for the same reasons as those found in the females-only results, although the differences in proportions of males in each phase were slightly less extreme than those of the females, resulting in fewer significant differences.

The most common phase was again variable, but not with the same patterning as found with the Meindl-Lovejoy results. Most collections did not have any phase I individuals; only Lisbon and Pretoria had individuals belonging to this phase. For Coimbra and Dart, phase V was the most common phase, while phase VII was most common for Grant individuals. Phase VI was most common for Lisbon and Spitalfields, while phase IV was most common for Pretoria. Peaks were not as pronounced for Buckberry-Chamberlain phase distributions, but Lisbon's phase distribution was slightly flatter and more evenly distributed than that of the other collections.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.323	.051	**.003**	**.000**	.393	.079	**.005**	**.000**	.351	.061
Dart			**.000**	**.000**	.773	.994	.250	.055	**.002**	**.000**
Grant					**.000**	**.000**	**.000**	**.000**	.042	.017
Lisbon							.171	.108	.029	**.001**
Pretoria									**.000**	**.000**

Table 4.61. Buckberry-Chamberlain, sexes pooled: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.782	.134	.010	**.000**	.813	.362	.048	**.003**	.713	.134
Dart			**.001**	**.000**	.900	.624	.518	.167	.051	**.004**
Grant					**.004**	**.000**	**.000**	**.000**	.053	**.003**
Lisbon							.100	.063	.288	.023
Pretoria									**.001**	**.000**

Table 4.62. Buckberry-Chamberlain, females only: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.760	.208	.123	.012	.225	.106	.191	.030	.880	.264
Dart			.006	**.000**	.687	.470	.336	.207	.093	.014
Grant					**.004**	**.000**	**.003**	**.000**	.314	.175
Lisbon							1.000	.671	.042	.011
Pretoria									.033	**.002**

Table 4.63. Buckberry-Chamberlain, males only: two-sample K-S and MWU results

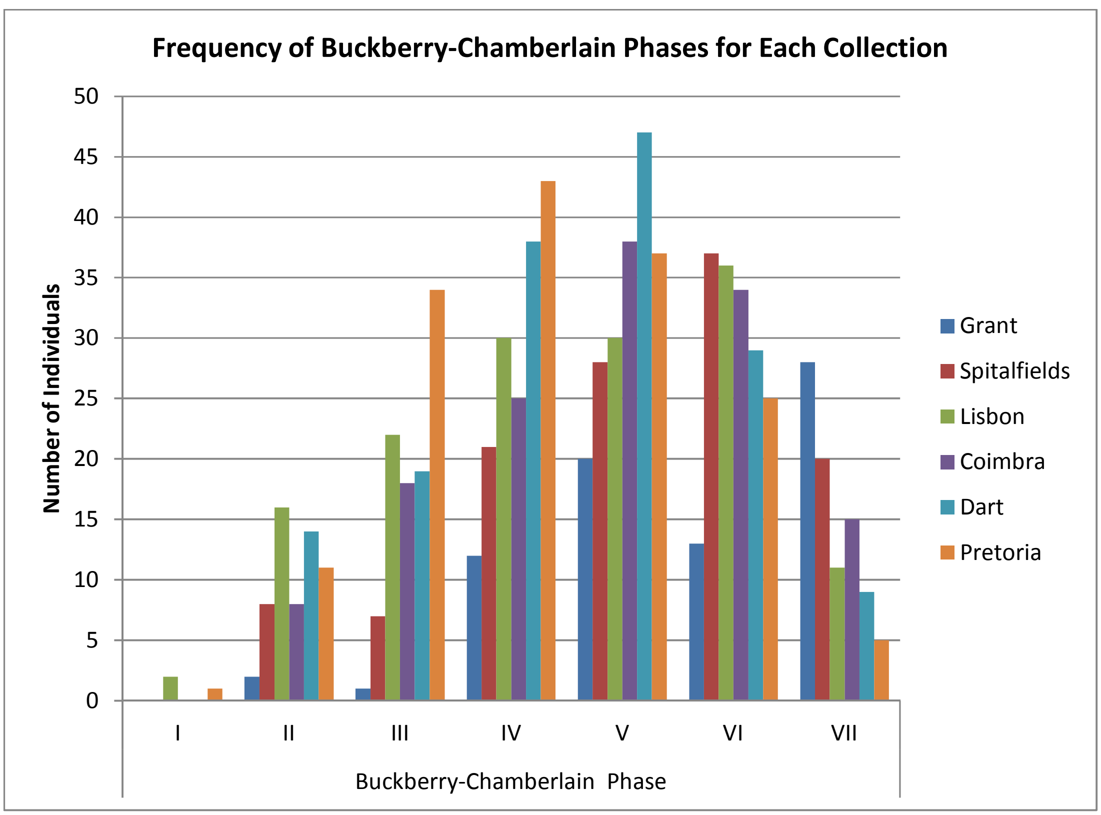


Figure 4.12. Bar chart of Buckberry-Chamberlain phase frequency for each collection, sexes pooled

	Phase																				
	I			II			III			IV			V			VI			VII		
	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**
Coimbra				3	5	**8**	11	7	**18**	13	12	**25**	18	20	**38**	19	15	**34**	8	7	**15**
Dart				7	7	**14**	14	5	**19**	14	24	**38**	20	27	**47**	13	16	**29**	7	2	**9**
Grant					2	**2**	1		**1**		12	**12**	2	18	**20**	4	9	**13**	9	19	**28**
Lisbon	1	1	**2**	10	6	**16**	10	12	**22**	10	20	**30**	17	13	**30**	21	15	**36**	6	5	**11**
Pretoria	1		**1**	6	5	**11**	15	19	**34**	23	20	**43**	18	19	**37**	9	16	**25**	3	2	**5**
Spitalfields				4	4	**8**	4	3	**7**	13	8	**21**	14	14	**28**	21	16	**37**	13	7	**20**

Table 4.64. Buckberry-Chamberlain phase frequency by collection
F: female; M: male; P: pooled.

4.8.4 Lateral-Anterior Cranial Suture Closure Method

Lateral-anterior cranial sutures also showed significant differences in phase distribution and median values for the pooled results (see Table 4.65, below). Spitalfields was found to be significantly different in median compared to all other collections except Coimbra, and in distribution compared to Lisbon; differences in distribution compared to all other collections except Coimbra neared significance. For females only (Table 4.66, below), differences nearing significance in distribution and median were found between Spitalfields compared to Grant and Lisbon, and in median between Spitalfields compared to Dart and Pretoria. Lisbon was significantly different in median compared to Pretoria and Spitalfields. For males only (see Table 4.67), the significant differences were again centred on Spitalfields; its median was significantly different to those of Dart, Grant, and Pretoria; its difference in distribution neared significance compared to Dart, Grant and Pretoria and its difference in median neared significance compared to Lisbon and Coimbra.

The phase frequencies for the pooled results (Table 4.68 and Figure 4.13) showed that the reason Spitalfields was significantly different from the other collections was the highly peaked frequency distribution (centred on phase 7); only phases 5, 6, 7 and 8 had any individuals at all. The other collections had relatively higher proportions of individuals in phases 5 and 6, further contributing to the distribution and median differences.

The Spitalfields female-only results followed the same highly-peaked phase frequency, centred on phase 7, leading to Spitalfields' different distribution and median compared to Lisbon and Grant. Spitalfields' female median is higher (nearing significance) than that of Dart and Pretoria females, due again to the highly peaked Spitalfields distribution focused on phase 7. Dart and Pretoria had flatter distributions, with more females in lower phases compared to Spitalfields. Lisbon's significantly lower median compared to Pretoria and nearing significance compared to Coimbra and Dart was because more Lisbon females were in the lower phases, and relatively fewer Lisbon females were in the higher phases (also giving a flatter distribution).

The phase frequencies for males only were less varied than those of females, following more closely the results for the sexes pooled. There were uniformly low numbers of males in phases 2, 3 and 4, with rising numbers of males in the phases that follow. Spitalfields was significantly different from the other collections due to the same peak in numbers of males in phase 7, and absolutely and relatively low proportions of males in phase 5 and 6 compared to the other collections.

While for all collections, phase 7 was the most common phase, the distributions for the other collections were flatter and more wide-ranging compared to Spitalfields. No collections had individuals in phase 1. Numbers of individuals in phases 2, 3 and 4 were generally quite low.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.842	.376	.784	.602	.354	.075	.546	.889	.152	.005
Dart			1.000	.837	.709	.267	.761	.369	.009	**.000**
Grant					.461	.281	1.000	.599	.019	**.002**
Lisbon							.051	.052	**.001**	**.000**
Pretoria									.022	**.001**

Table 4.65. Lateral-anterior sutures, sexes pooled: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.885	.464	.154	.223	.084	.009	.825	.634	.487	.103
Dart			.496	.492	.362	.031	.996	.712	.203	.012
Grant					.227	.501	.536	.323	.040	.010
Lisbon							.059	.011	.007	**.000**
Pretoria									.171	.019

Table 4.66. Lateral-anterior sutures, females only: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.519	.986	.586	1.000	.909	.789	.865	.132	.006
Dart			1.000	.994	1.000	.450	.930	.344	.026	**.000**
Grant					.997	.517	.970	.386	.021	**.001**
Lisbon							.921	.934	.105	.008
Pretoria									.094	**.002**

Table 4.67. Lateral-anterior sutures, males only: two-sample K-S and MWU results

4.8.5 Vault Cranial Suture Closure Method

The vault suture results for the sexes pooled also showed some significant differences in phase distribution and median, but with less obvious patterning than that of the lateral-anterior sutures (see Table 4.69 for vault suture results). There were significant differences in both distribution and median between Spitalfields and Dart. Further significant differences in median only were found between Grant compared to Dart and Pretoria, and between Pretoria and Spitalfields. Other differences neared significance. Grant did not have enough females with available vault sutures to compare to the females of the other collections; as such, the pooled results for Grant were mostly based on male values. For females only, differences in median between Dart compared to Coimbra and Spitalfields neared significance (see Table A5.2, Appendix

	Phase																				
	2			**3**			**4**			**5**			**6**			**7**			**8**		
	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**
Coimbra	1	0	**1**	3	1	**4**	1	2	**3**	9	8	**17**	19	13	**32**	40	40	**80**	0	2	**2**
Dart	0	0	**0**	1	1	**2**	0	1	**1**	12	9	**21**	26	21	**47**	32	44	**76**	0	0	**0**
Grant	0	0	**0**	0	1	**1**	0	0	**0**	0	7	**7**	9	17	**26**	2	29	**31**	0	2	**2**
Lisbon	0	0	**0**	7	0	**7**	3	2	**5**	14	8	**22**	23	15	**38**	24	39	**63**	0	3	**3**
Pretoria	0	0	**0**	0	0	**0**	0	0	**0**	8	4	**12**	31	23	**54**	31	45	**76**	0	0	**0**
Spitalfields	0	0	**0**	0	0	**0**	0	0	**0**	2	0	**2**	16	4	**20**	33	28	**61**	0	2	**2**

Table 4.68. Lateral-anterior suture phase frequency by collection
F: female; M: male; P: pooled.

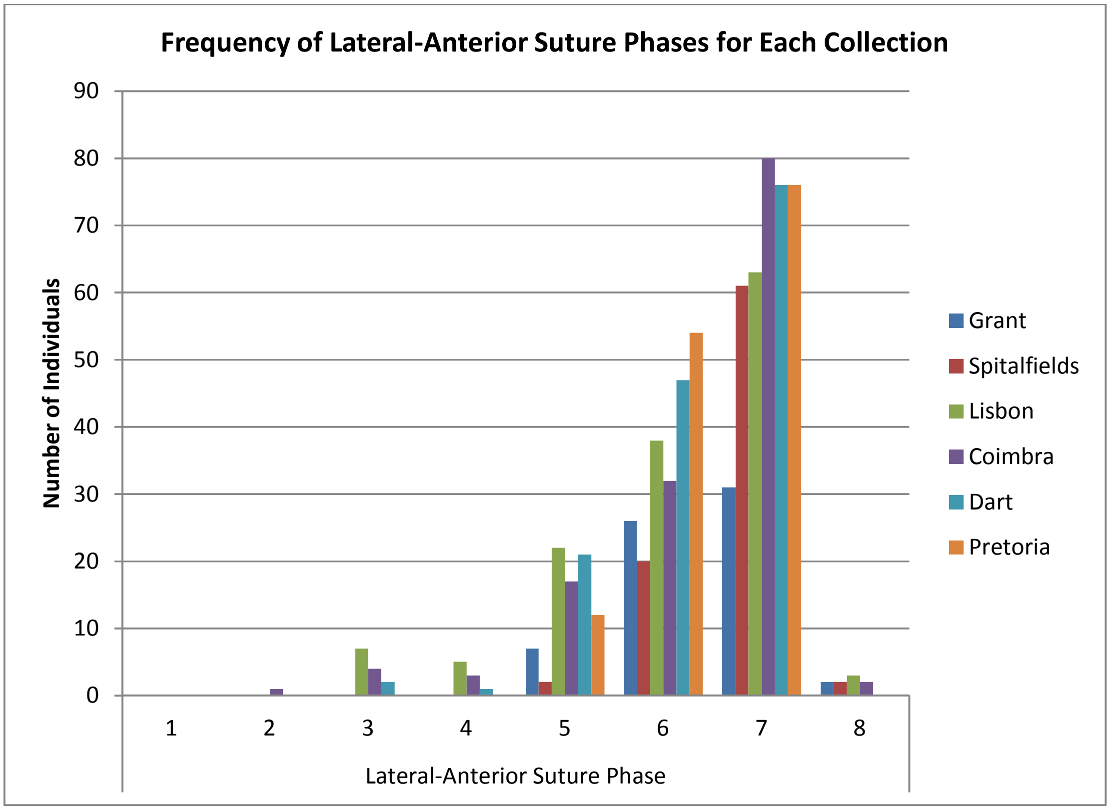


Figure 4.13. Bar chart of lateral-anterior suture phase frequency for each collection, sexes pooled

5). For males only (Table 4.70), significant differences were present in the distribution and median between Pretoria and Spitalfields, and median only between Grant and Pretoria. Further differences neared significance in phase distribution and median were found between Spitalfields and the South African collections, and in median only between Coimbra and Pretoria males.

The phase frequencies are presented below (in Table 4.71 and Figure 4.14). The results for the sexes pooled showed that part of the reason for Grant's significant differences (mostly in median) compared to the other collections was the low numbers of individuals in the lower phases, and relatively high proportions of individuals in the highest phases. However, the small sample size of Grant individuals with vault sutures available for study likely contributed to the significant differences. The differences in distribution and median between Spitalfields and the South African collections were due to Spitalfields' fairly steeply-peaked distribution and high proportion of individuals in phase 5. In contrast, the highest proportion of individuals for both Dart and Pretoria was in phase 4, with the distribution rising quickly and flattening out somewhat at the higher phases compared to Spitalfields. This was also the reason for Spitalfields' higher median value compared to Dart and Pretoria. Dart's median was lower than Coimbra's, again due to the majority of Dart individuals being in phase 4; Coimbra's highest proportion of individuals was in phase 5, and the distribution was somewhat flatter.

When females were considered alone, the differences in median of Dart females compared to both Coimbra and Spitalfields were due to the same proportional differences – that is, Dart's lower median was because of the higher proportion of females in phase 4 compared to Coimbra and Spitalfields.

For males alone, the Grant differences in median compared to Coimbra, Dart, Lisbon and Pretoria (and distribution with Pretoria), were because of the same reasons described for the pooled results above; low sample size and higher proportions of Grant individuals in the higher phases. Spitalfields males were different in phase distribution and median compared to the South African collections because of Spitalfields' relatively higher proportions of individuals in the older age phases and low proportions of individuals in the lower phases. Coimbra and Pretoria differ in median for similar reasons; Coimbra had a higher median value as a higher proportion of males were in the higher phases compared to Pretoria.

Overall, no individuals were in phase 1, and relatively few individuals were in phases 2 and 3. The most common phase for Pretoria and Dart individuals was phase 4, while for Coimbra, Lisbon, and Spitalfields, the most common phase was 5. For Grant, the most common phase was 7.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.067	.038	.205	.011	.998	.581	.302	.061	.682	.173
Dart			.058	**.001**	.400	.110	.980	.670	**.002**	**.001**
Grant					.154	.005	.037	**.001**	.304	.037
Lisbon							.683	.181	.217	.053
Pretoria									.016	**.001**

Table 4.69. Vault sutures, sexes pooled: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.620	.291	.193	.016	.997	.761	.160	.014	.401	.113
Dart			.108	.006	1.000	.503	.626	.190	.018	.013
Grant					.222	.016	.014	**.001**	.426	.072
Lisbon							.256	.052	.098	.080
Pretoria									**.002**	**.000**

Table 4.70. Vault sutures, males only: two-sample K-S and MWU results

	Phase																	
	2			3			4			5			6			7		
	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**
Coimbra	2	0	**2**	10	5	**15**	19	13	**32**	22	26	**48**	10	16	**26**	10	6	**16**
Dart	4	2	**6**	9	3	**12**	31	25	**56**	14	23	**37**	9	15	**24**	4	7	**11**
Grant	0	0	**0**	0	0	**0**	0	1	**1**	1	2	**3**	0	3	**3**	0	4	**4**
Lisbon	2	0	**2**	5	5	**10**	25	17	**42**	23	20	**43**	12	15	**27**	4	7	**11**
Pretoria	0	0	**0**	3	6	**9**	27	27	**54**	19	26	**45**	12	8	**20**	2	4	**6**
Spitalfields	0	0	**0**	3	1	**4**	15	2	**17**	18	17	**35**	3	12	**15**	9	3	**12**

Table 4.71. Vault suture phase frequency by collection
F: female; M: male; P: pooled.

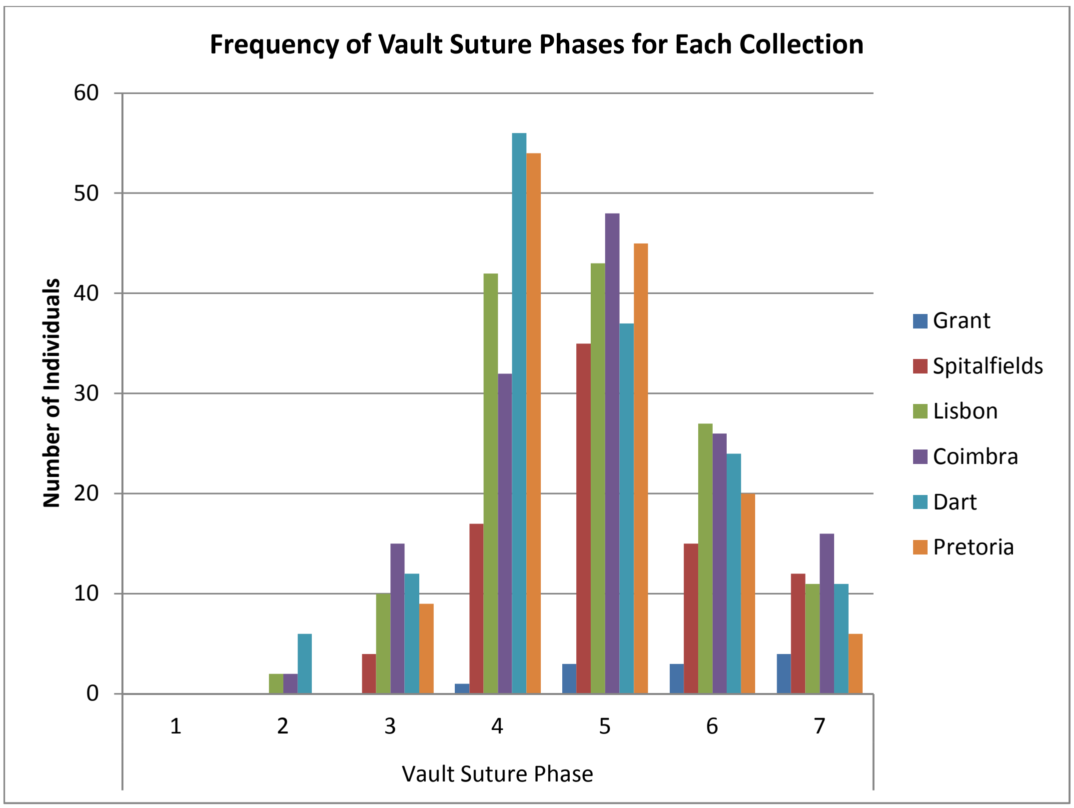


Figure 4.14. Bar chart of vault suture phase frequency for each collection, sexes pooled

4.8.6 Sternal End of Fourth Rib Method

For the fourth rib, sample sizes were fairly small, ranging from two individuals from the Grant Collection to 65 individuals from the Pretoria Collection. Because individuals were then placed in fourth rib phases 3 to 15, sample sizes for any one phase were very small, making meaningful analysis difficult. Regardless, an attempt was made to examine distribution and median differences. The pooled results showed only one instance of a difference approaching significance, in the phase distributions of the Lisbon and Dart collections (Table A5.3, Appendix 5). The female- and male-only results showed no significant differences in phase distribution or median (see Tables A5.4 and A5.5, Appendix 5), but the splitting of the sexes made sample sizes even smaller, again making meaningful analysis difficult. The phase frequencies (fully presented in Appendix 5, Table A5.6a and b) indicated that the Lisbon and Dart differences lie in Dart's high proportion of individuals in phase 11 (16 individuals compared to Lisbon's 5 in this phase), and Lisbon's higher proportions of individuals in phases 8 to 10. However, the importance of these results may be limited due to the small numbers of individuals in each phase.

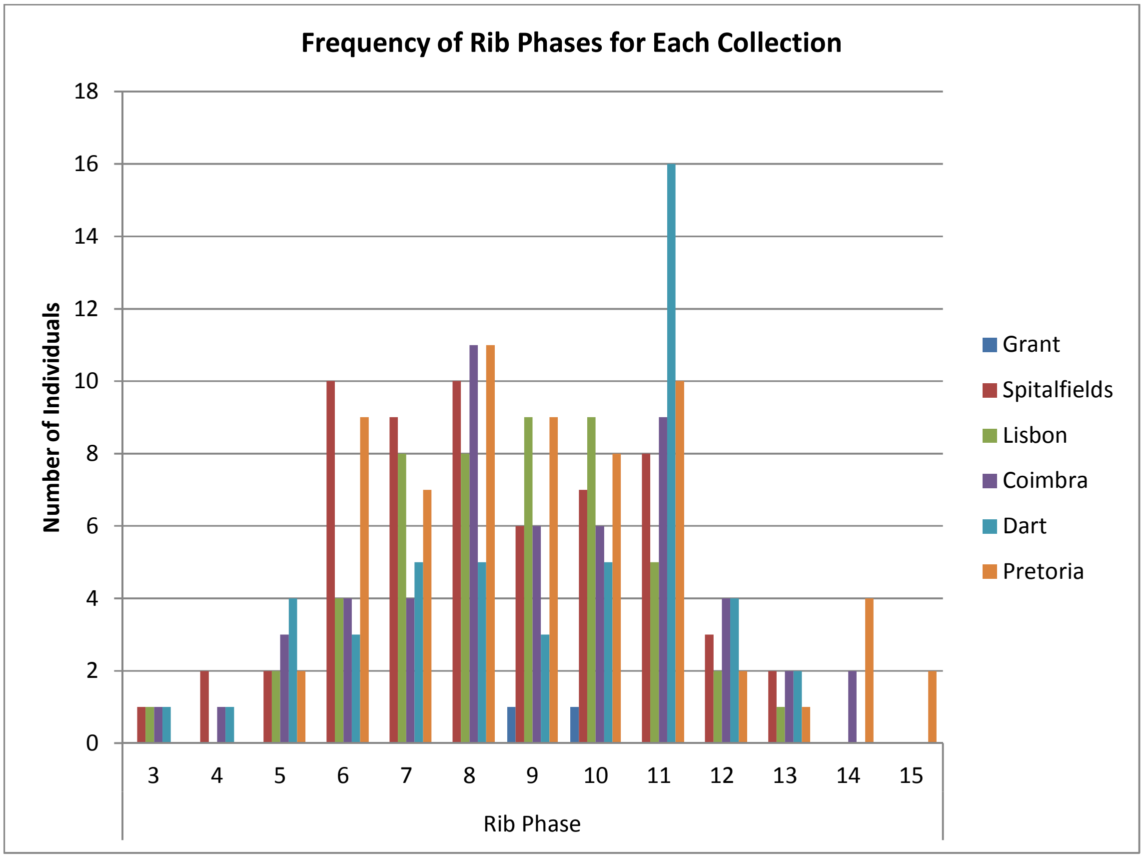


Figure 4.15. Bar chart of fourth rib phase frequency for each collection, sexes pooled

4.9 Buckberry-Chamberlain Auricular Surface Characteristics – Variation in Score Distribution

As the Buckberry-Chamberlain auricular surface method combines scored traits, it was possible to analyse the scored traits separately for median and distribution variation. The distribution and medians for scores of transverse organisation will be discussed first, followed by surface texture, microporosity, macroporosity, and finally, apical changes.

4.9.1 Transverse Organisation

Significant differences were found in both score distribution and median for transverse organisation scores when the sexes were pooled. These differences were found between the Grant Collection compared to Lisbon and Pretoria, and in median only between Grant compared to Coimbra and Dart. Differences in distribution nearing significance were found in Grant compared to Coimbra and Dart. Pretoria was found to have significant distribution and median differences compared to Spitalfields, and in median only in Lisbon compared to Spitalfields (see Table 4.72 for details). Other differences nearing significance were in distribution and median between Pretoria and Dart, and Pretoria and Coimbra. The females-only results revealed largely the same significant differences (see Table 4.73). The results for males only (Table 4.74), however, showed fewer significant differences; only the medians were significantly different in Grant compared to Lisbon and Pretoria, and nearing significance between Dart and Lisbon, and Coimbra and Grant.

The score frequencies (presented in Table 4.75 and Figure 4.16) showed that Grant had lower proportions of individuals with scores of 1 and 2 and higher proportions of 4, and especially 5, than the other collections, resulting in the significant differences in distribution. Spitalfields shared these characteristics with Grant, but to a less extreme degree, hence the significant differences with only Lisbon and Pretoria. Pretoria had higher proportions of scores of 2 and 3 and lower proportions of scores of 4 and 5 compared to Coimbra and Dart, resulting in the significant differences in distribution and median. Lisbon's significant differences in median compared to Coimbra and Dart were for the same reasons, but were confined to median as the proportions were not as extreme as in Pretoria.

The significant differences in the females-only results were because of the same differences in proportion as for the pooled results, described above. For males only, the Grant median was significantly higher than that of Lisbon, Coimbra and Pretoria, again due to the high proportion of individuals with scores of 4 and 5. The significant difference in median between Lisbon and Dart was due to Lisbon's relatively higher proportions of lower scores.

In general, scores of 4 were the most common for all collections except Pretoria, for which scores of 3 were more common. For Coimbra, Dart and Lisbon, scores of 3 were the next most common, while for Pretoria, the second most common score was 4. For Spitalfields and Grant, the next most common score was 5.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.986	.850	.029	**.000**	.193	.033	.033	.032	.681	.238
Dart			.021	**.000**	.265	.012	.047	.010	.594	.277
Grant					**.003**	**.000**	**.000**	**.000**	.174	.026
Lisbon							.999	.985	.025	**.003**
Pretoria									**.003**	**.003**

Table 4.72. Buckberry-Chamberlain auricular surface, sexes pooled, transverse organisation score: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.928	.325	.123	**.001**	.232	.032	.010	.007	.971	.422
Dart			.021	**.000**	.970	.207	.182	.048	.536	.092
Grant					**.003**	**.000**	**.000**	**.000**	.260	.012
Lisbon							.843	.447	.072	.007
Pretoria									**.002**	**.002**

Table 4.73. Buckberry-Chamberlain auricular surface, females only, transverse organisation score: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.828	.217	.211	.009	.976	.395	.998	.795	.952	.421
Dart			.375	.102	.400	.029	.528	.100	.755	.890
Grant					.079	**.001**	.157	**.003**	.228	.137
Lisbon							.834	.485	.567	.140
Pretoria									.694	.293

Table 4.74. Buckberry-Chamberlain auricular surface, males only, transverse organisation score: two -sample K-S and MWU results

	Score														
	1			**2**			**3**			**4**			**5**		
	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**
Coimbra	1	0	**1**	11	14	**25**	16	15	**31**	36	31	**67**	8	5	**13**
Dart	0	0	**0**	12	9	**21**	23	22	**45**	33	41	**74**	5	9	**14**
Grant	1	2	**3**	0	0	**0**	0	19	**19**	8	23	**31**	7	16	**23**
Lisbon	0	0	**0**	18	18	**36**	24	19	**43**	30	28	**58**	3	5	**8**
Pretoria	0	0	**0**	19	11	**30**	28	26	**54**	18	30	**48**	6	5	**11**
Spitalfields	3	2	**5**	9	10	**19**	11	8	**19**	31	23	**54**	13	9	**22**

Table 4.75. Transverse organisation score frequency by collection
F: female; M: male; P: pooled.

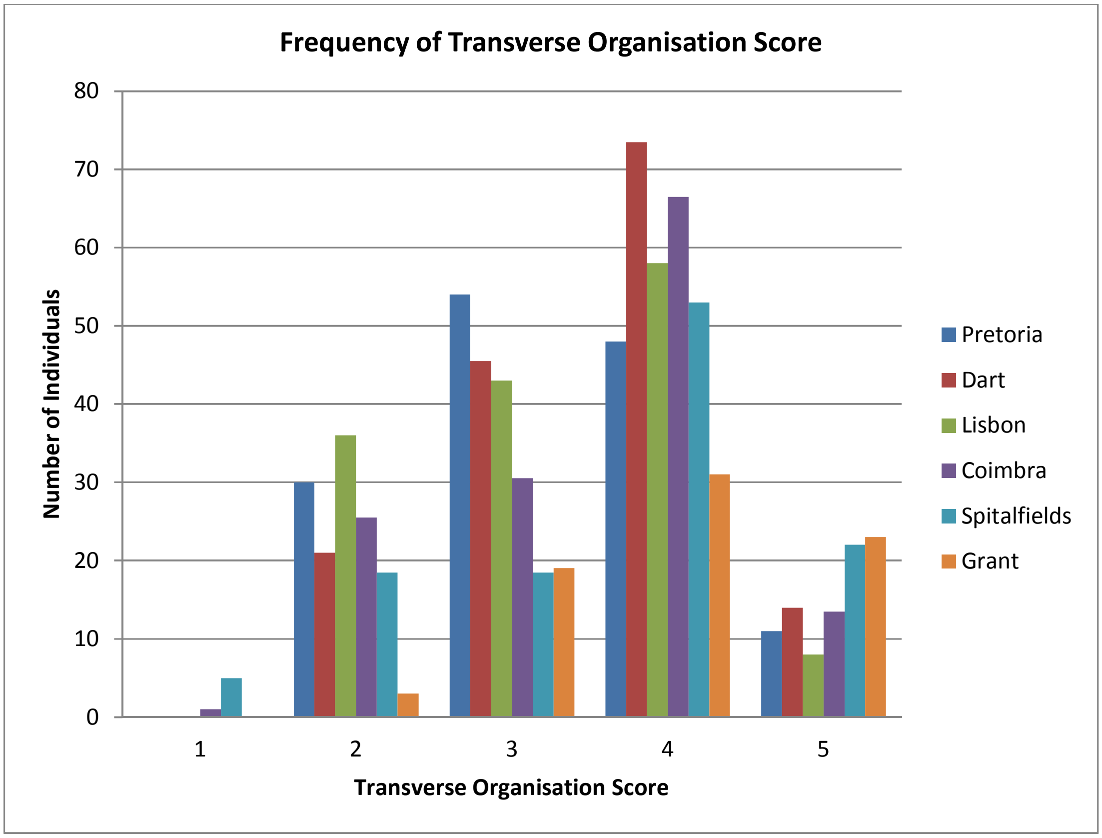


Figure 4.16. Bar chart of Buckberry-Chamberlain transverse organisation score frequency for each collection, sexes pooled

4.9.2 Surface Texture

For the pooled surface texture results (Table 4.76), the significant differences were again centred on Grant and Spitalfields. Both collections showed significant differences in distribution and median with all other collections except each other, where the difference was in median only. The females-only results were similar (Table 4.77, below): Grant females were significantly different in score distribution and median to all other collections except Spitalfields, where the difference was in median only. Spitalfields females were significantly different in distribution and median to Pretoria females, and in median only to the Portuguese collections and Dart. For males only (Table 4.78), both Grant and Spitalfields were significantly different in distribution and median to nearly all other collections, with the exception of each other and the distribution between Spitalfields and Coimbra, which neared significance.

As with transverse organisation, the score frequencies for Grant and Spitalfields had higher proportions of higher scores and lower proportions of scores of 1 to 3 than the other collections (see Table 4.79 and Figure 4.17 for score frequencies). The median alone was significantly different between Spitalfields and Grant despite the similarity in their score distributions, because

Grant had higher proportions of scores of 4 and, in particular, 5. These reasons hold for the significant differences found in the female- and male-only results.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.167	.981	**.000**	**.000**	1.00	.703	.101	.563	**.005**	**.000**
Dart			**.000**	**.000**	.286	.660	1.00	.514	**.000**	**.000**
Grant					**.000**	**.000**	**.000**	**.000**	.155	**.001**
Lisbon							.120	.310	**.004**	**.000**
Pretoria									**.000**	**.000**

Table 4.76. Buckberry-Chamberlain auricular surface, sexes pooled, surface texture score: two - sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.703	.974	**.002**	**.000**	1.00	.742	.386	.733	.303	**.004**
Dart			**.000**	**.000**	.473	.674	1.00	.706	.010	**.002**
Grant					**.003**	**.000**	**.000**	**.000**	.079	**.001**
Lisbon							.188	.402	.423	.009
Pretoria									**.002**	**.000**

Table 4.77. Buckberry-Chamberlain auricular surface, females only, surface texture score: two - sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.455	.921	**.002**	**.000**	1.00	.838	.492	.633	.018	**.002**
Dart			**.000**	**.000**	.798	.853	1.00	.601	**.000**	**.000**
Grant					**.000**	**.000**	**.000**	**.000**	.793	.088
Lisbon							.879	.543	**.004**	**.004**
Pretoria									**.000**	**.000**

Table 4.78. Buckberry-Chamberlain auricular surface, males only, surface texture score: two - sample K-S and MWU results

	Score														
	1			2			3			4			5		
	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**	F	M	**P**
Coimbra	4	7	**11**	15	10	**25**	23	21	**44**	30	30	**60**	1	1	**2**
Dart	1	2	**3**	10	7	**17**	41	48	**89**	21	25	**46**	3	2	**5**
Grant	0	0	**0**	1	2	**3**	0	11	**11**	7	35	**42**	8	12	**20**
Lisbon	5	3	**8**	12	12	**24**	24	27	**51**	34	25	**59**	0	4	**4**
Pretoria	1	1	**2**	9	11	**20**	45	43	**88**	20	25	**45**	1	0	**1**
Spitalfields	2	1	**3**	5	7	**12**	21	6	**27**	30	34	**64**	10	4	**14**

Table 4.79. Surface texture score frequency by collection
F: female; M: male; P: pooled.

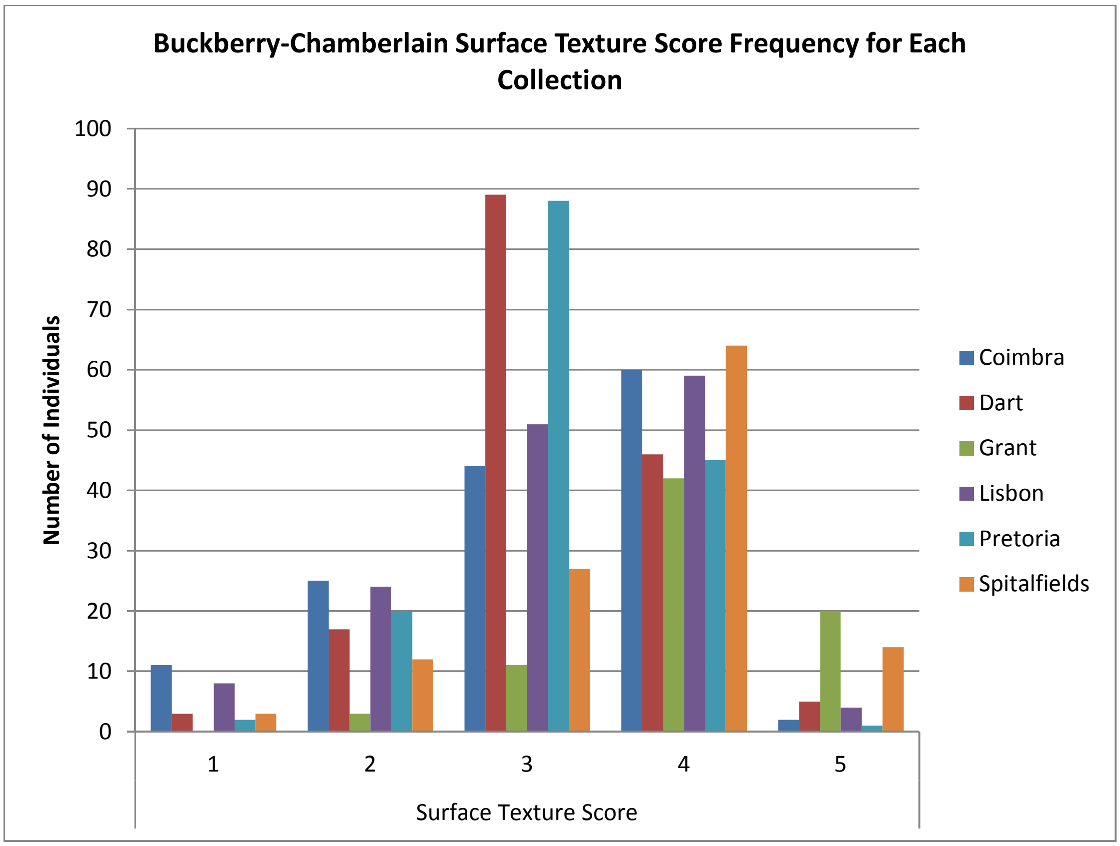


Figure 4.17. Bar chart of Buckberry-Chamberlain surface texture score frequency for each collection, sexes pooled

4.9.3 Microporosity

The significant differences in microporosity for the pooled results (Table 4.80) were found in score distribution and median for Dart compared to Coimbra and Spitalfields, and in Pretoria compared to Coimbra, Grant, Lisbon and Spitalfields. Significant differences in median only were found in Lisbon compared to Coimbra, Dart and Spitalfields and between Dart and Grant. For females only (Table 4.81), there were fewer significant differences, although where they did exist, they were in the same combinations as with the pooled results: in score distribution and median between Pretoria and Spitalfields, and in median only between Coimbra compared to Dart and Pretoria, and between Dart and Spitalfields. For males only (Table 4.82), significant differences in distribution and median were found in Pretoria compared to Coimbra, Grant and Spitalfields and between Dart and Coimbra. Significant differences in median only were found between Lisbon and Coimbra, and between Dart and Spitalfields. Other differences neared significance, as in distribution and median between Dart and Grant.

The microporosity score frequencies are presented in Table 4.83 and Figure 4.18, below. The pooled results show that the South African collections had flatter score distributions, with higher proportions of scores of 1 and 2 and lower proportions of scores of 3 than the other collections, explaining the significant differences compared to the other collections. While Lisbon followed the opposite pattern, like Grant, Coimbra, and Spitalfields, with lower proportions of individuals with scores of 1 and 2 and very high proportions of individuals with scores of 3, it was less extreme, so only the median is significantly different between Lisbon and the South African collections. This was also the reason for the significant difference in median between Lisbon and Coimbra; Coimbra's higher proportion of scores of 3 resulted in a higher median.

The females-only results followed the same patterns, but absolute and relative proportions of Dart and Pretoria females were slightly less extreme than those of the pooled results, leading to fewer significant differences. In terms of the males-only results, the fewer high-scoring South African males resulted in more significant differences in males (between the South African collections and other collections) compared to females. The reasons remain the same. The male differences in median in Lisbon compared to Spitalfields and Coimbra were also for the same reasons as in the pooled results.

In general, scores of 3 were the most common for all collections; however, Pretoria and Dart had slightly flatter, more evenly-distributed score frequencies compared to the other collections.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	**.000**	**.000**	.798	.078	.122	**.004**	**.000**	**.000**	1.00	.996
Dart			.042	**.006**	.082	**.006**	.875	.592	**.000**	**.000**
Grant					1.00	.544	**.002**	**.001**	.821	.091
Lisbon							**.003**	**.001**	.150	**.007**
Pretoria									**.000**	**.000**

Table 4.80. Buckberry-Chamberlain auricular surface, sexes pooled, microporosity score: two - sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.068	**.002**	1.00	.401	.992	.262	.011	**.000**	1.00	.533
Dart			.985	.276	.418	.039	1.00	.797	.021	**.000**
Grant					1.00	.896	.789	.188	.980	.212
Lisbon							.109	.014	.775	.089
Pretoria									**.003**	**.000**

Table 4.81. Buckberry-Chamberlain auricular surface, females only, microporosity score: two - sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	**.000**	**.000**	.811	.097	.085	**.004**	**.000**	**.000**	1.00	.501
Dart			.036	.007	.462	.085	.935	.582	.005	**.000**
Grant					.887	.271	**.002**	**.001**	1.00	.364
Lisbon							.054	.020	.361	.044
Pretoria									**.000**	**.000**

Table 4.82. Buckberry-Chamberlain auricular surface, males only, microporosity score: two - sample K-S and MWU results

	Score								
	1			**2**			**3**		
	F	M	**P**	F	M	**P**	F	M	**P**
Coimbra	1	2	**3**	15	9	**24**	56	54	**110**
Dart	13	16	**29**	19	27	**46**	41	39	**80**
Grant	1	7	**8**	4	10	**14**	11	43	**54**
Lisbon	5	9	**14**	17	18	**35**	53	43	**96**
Pretoria	9	14	**23**	26	33	**59**	36	28	**64**
Spitalfields	0	3	**3**	12	8	**20**	54	40	**94**

Table 4.83. Microporosity score frequency by collection
F: female; M: male; P: pooled.

4.9.4 Macroporosity

The patterning of differences for macroporosity was less obvious than that of microporosity. The pooled results (Table 4.84) showed differences nearing significance in score distribution and median between Pretoria and Coimbra, and in median between Coimbra compared to Dart and Lisbon; values approaching significance also occurred between Pretoria compared to Coimbra and Spitalfields, and in Lisbon compared to Coimbra and Spitalfields. Also nearing significance were differences in median between Grant compared to Dart, Lisbon and Pretoria. Fewer

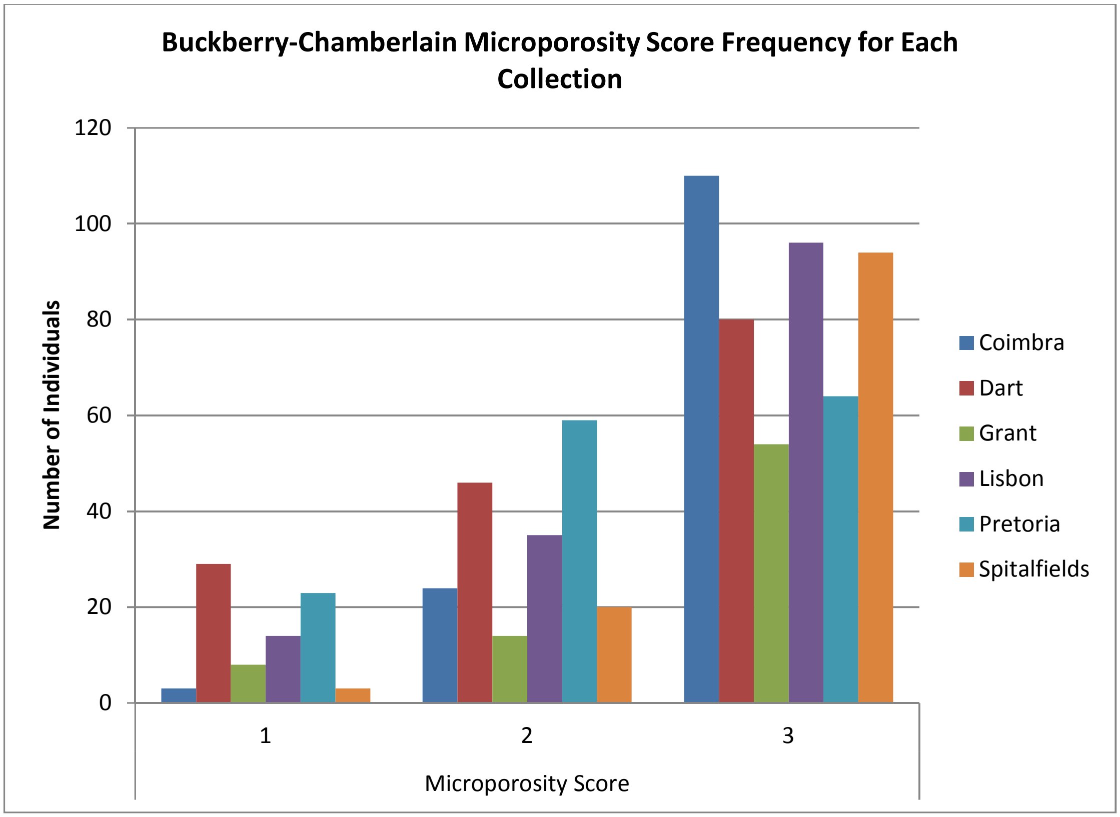


Figure 4.18. Bar chart of Buckberry-Chamberlain microporosity score frequency for each collection, sexes pooled

differences were found in the females-only results (Table A5.7, Appendix 5); score distribution and median were significantly different in Coimbra compared to Lisbon and Pretoria, and in median only in Spitalfields compared to Lisbon and Pretoria. Fewer differences still were found in the males-only results (Table A5.8, Appendix 5); medians only were significantly different between Dart and Coimbra, and in Grant compared to Dart and Pretoria.

Score frequencies are presented in Table 4.85 and Figure 4.19. Coimbra and Spitalfields had relatively lower proportions of scores of 1 and relatively higher proportions of scores of 2 compared to the other collections, resulting in the differences in distribution and median. This also accounted for the females-only difference between Coimbra compared to Pretoria and Lisbon in distribution and median, and the median only differences in Spitalfields compared to Lisbon and Pretoria.

For males only, Dart's lower median compared to Grant and Coimbra was because of Dart's higher proportion of scores of 1 and lower proportion of scores of 2. Similarly, Pretoria's higher proportion of scores of 1 and lower proportion of scores of 2 compared to Grant resulted in Pretoria's lower median.

In general, scores of 3 were least common. For Coimbra and Spitalfields, scores of 2 were the most common, while scores of 1 were the most common for Dart, Grant, Lisbon and Pretoria. The score distributions for Grant, Coimbra and Spitalfields were flatter and more evenly spread than those of Dart, Lisbon, and Pretoria.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	**.004**	.007	.407	.727	**.004**	.005	**.002**	**.002**	1.00	.780
Dart			.246	.022	1.00	.895	1.00	.707	.036	.025
Grant					.255	.016	.165	.010	.624	.596
Lisbon							1.00	.806	.040	.017
Pretoria									.020	.009

Table 4.84. Buckberry-Chamberlain auricular surface, sexes pooled, macroporosity score: two - sample K-S and MWU results

	Score								
	1			**2**			**3**		
	F	M	**P**	F	M	**P**	F	M	**P**
Coimbra	21	27	**48**	38	31	**69**	13	7	**20**
Dart	36	50	**86**	19	24	**43**	18	7	**25**
Grant	6	26	**32**	3	21	**24**	7	14	**21**
Lisbon	41	40	**81**	21	22	**43**	13	8	**21**
Pretoria	38	44	**82**	22	19	**41**	11	9	**20**
Spitalfields	22	23	**45**	29	24	**53**	15	4	**19**

Table 4.85. Macroporosity score frequency by collection
F: female; M: male; P: pooled.

4.9.5 Apical Change

For apical change scores, the pooled results (Table 4.86) showed significant differences in score distribution and median in Grant compared to Dart and Pretoria and in median between Grant and all other collections, with the differences in distribution also nearing significance. Significant differences in median only were seen between Pretoria and Spitalfields, and nearing significance between Pretoria compared to Coimbra, Dart and Lisbon. When females were considered alone (Table 4.87), significant differences in median were found between Pretoria compared to Grant, Lisbon and Spitalfields (and differences in distribution in the same approaching significance). Differences approaching significance in median only were between Grant compared to Coimbra, Dart, Lisbon and Spitalfields, as well as between Lisbon and Dart, Coimbra and Pretoria, and Spitalfields compared to Dart and Grant. Fewer significant differences were found when males were considered alone (Table 4.88), as they occurred in median only between Grant and Lisbon and Grant and Pretoria; differences in median between Grant and every other collection except Spitalfields neared significance.

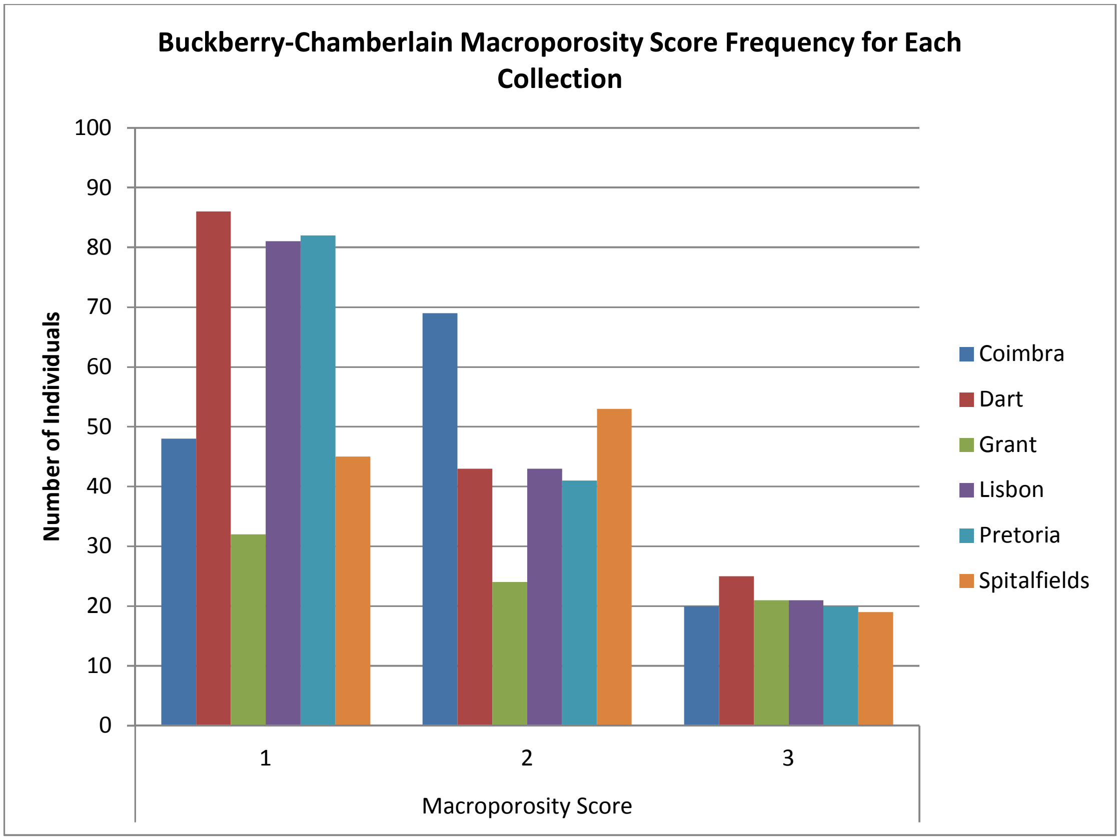


Figure 4.19. Bar chart of Buckberry-Chamberlain macroporosity score frequency for each collection, sexes pooled

The score frequencies (Table 4.89 and Figure 4.20) revealed that the differences between the Grant pooled results and the other collections lie in Grant's low proportion of scores of 1, somewhat low proportion of scores of 2, and high proportion of scores of 3 compared to the other collections. While Spitalfields followed the same pattern as Grant, the difference was not as extreme as with Grant, only resulting in a significant difference in median with Pretoria, with its opposite pattern of a high proportion of scores of 1, and somewhat low proportion of scores of 3 (compared to Spitalfields). Pretoria's score distribution resulted in a lower median than that of Coimbra, Dart, and Lisbon.

For females alone, Grant again had high proportions of scores of 2 and 3, leading to the significant differences in median compared to the other collections. Pretoria again had the opposite pattern, resulting in the difference in distribution with Grant. Spitalfields and Lisbon (with relatively lower proportions of scores of 1 and high proportions of scores of 3) were different from Pretoria for the same reasons. These proportions for Pretoria also resulted in a lower median compared to Coimbra (with its higher proportions of scores of 2 and 3). Dart's median was significantly lower than that of Lisbon and Spitalfields due to its relatively high proportion of scores of 1, and low proportion of scores of 3.

The males-only differences, in Grant's median compared to all collections except Spitalfields, again were due to the high proportion of Grant individuals with scores of 2, and particularly 3, and a low proportion of scores of 1. Spitalfields males followed the Grant pattern, but with a slightly lower proportion of scores of 3, so that differences were not significant.

While scores of 2 were the most common score across all collections, the Grant distribution was more "top-heavy" (more scores of 2 and 3), while that of Pretoria is slightly "bottom-heavy" (more scores of 1 and 2). The other collections displayed distributions that were closer to a normal distribution, with scores of 2 as the peak.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.0	.708	.005	**.000**	.990	.730	.179	.018	.998	.252
Dart			**.004**	**.000**	.988	.483	.427	.045	.811	.134
Grant					.043	**.000**	**.000**	**.000**	.045	**.003**
Lisbon							.370	.010	.887	.452
Pretoria									.030	**.001**

Table 4.86. Buckberry-Chamberlain auricular surface, sexes pooled, apical change score: two - sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.814	.181	.138	.005	.738	.375	.256	.012	.917	.213
Dart			.063	**.001**	.326	.038	.994	.269	.124	.011
Grant					.483	.039	.015	**.000**	.294	.026
Lisbon							.045	**.002**	.900	.812
Pretoria									.016	**.000**

Table 4.87. Buckberry-Chamberlain auricular surface, females only, apical change score: two - sample K-S results and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.504	.098	.013	1.000	.594	.890	.385	1.000	.675
Dart			.213	.043	.978	.225	.540	.124	1.000	.873
Grant					.064	**.004**	.086	**.002**	.389	.061
Lisbon							1.000	.723	1.00	.375
Pretoria									.921	.243

Table 4.88. Buckberry-Chamberlain auricular surface, males only, apical change score: two - sample K-S and MWU results

	Score								
	1			**2**			**3**		
	F	M	**P**	F	M	**P**	F	M	**P**
Coimbra	21	12	**33**	38	41	**79**	13	14	**27**
Dart	29	12	**41**	34	49	**83**	10	20	**30**
Grant	1	7	**8**	7	28	**35**	8	26	**34**
Lisbon	22	16	**38**	31	41	**72**	22	14	**36**
Pretoria	34	21	**55**	35	39	**74**	5	16	**21**
Spitalfields	13	9	**22**	38	29	**67**	15	13	**28**

Table 4.89. Apical change score frequency by collection
F: female; M: male; P: pooled.

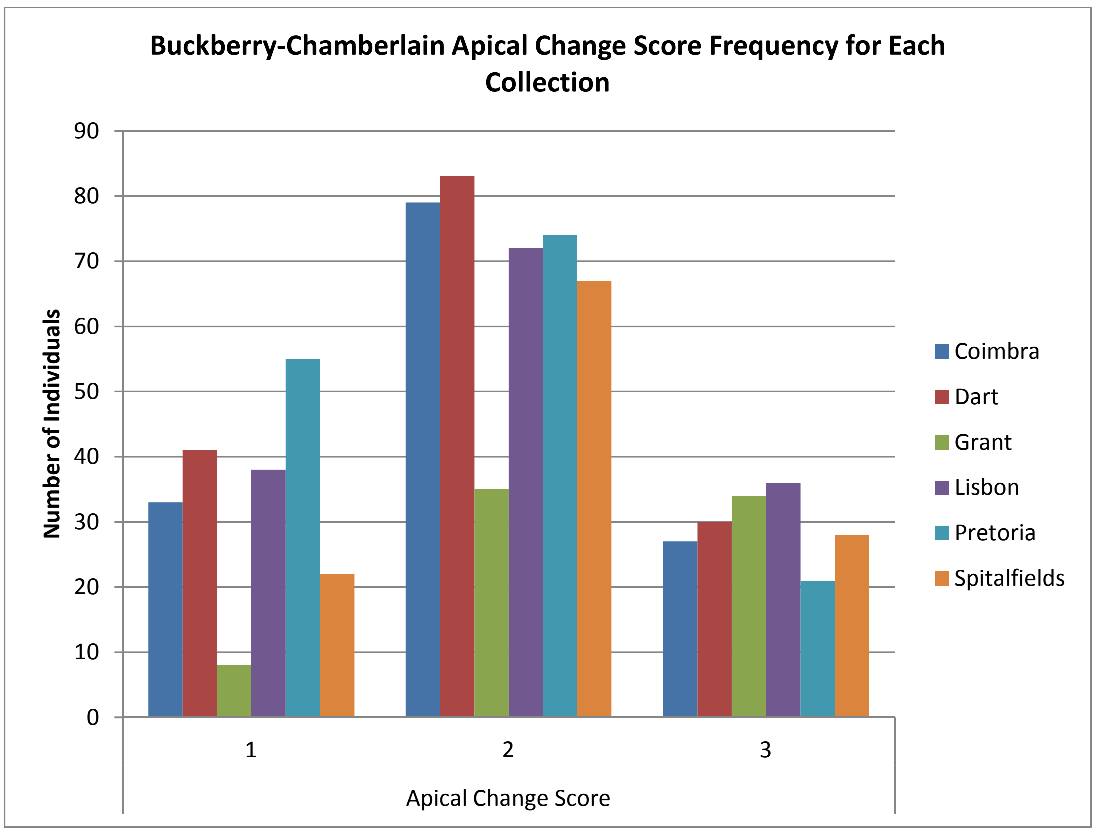


Figure 4.20. Bar chart of Buckberry-Chamberlain apical change score frequency for each collection, sexes pooled

4.10 Within-Collection Distribution Variation by Sex

For each method, the phase or score distribution was compared by sex for each collection, to look for within-collection sex differences. Each method had at least one collection with significant differences between males and females, although these were not always in both median and distribution. For example, only Coimbra showed significant differences in phase distribution and median for the Suchey-Brooks method, and these differences were due to lower proportions of females in phases III and V and higher proportions in phases IV and VI compared to males. Although the fourth rib had significant differences between the sexes in all collections tested, small sample sizes likely played a role. As no method displayed consistent sex differences across collections, the detailed results are presented in Table A5.9, Appendix 5.

Within-collection sex differences in score distributions were also analysed for each of the Buckberry-Chamberlain auricular surface skeletal traits that are scored individually. Fewer sex differences were found in these traits compared to the phase distributions for each method. For example, a significant difference in score distribution and median were found between Dart males and females for apical change. This was because there was a higher proportion of Dart females with apical change scores of 1 and lower proportions of scores of 2 and 3, resulting in a significant distribution difference and higher male median. As differences were few and not consistent across all collections, further details are provided in Table A5.10, Appendix 5.

4.11 Variation in Phase/Score Distribution by Age Group

The data were further subdivided into age groups (ten-year age groups by decade; e.g. 20 to 29 years, 30 to 39 years, etc.), to explore whether variation occurred within specific age ranges, using the K-S and MWU tests. As only the Dart Collection had any individuals in the 100 and over age group, phase distributions for this age group were not tested.

No specific age-related trends were found. For the Meindl-Lovejoy and Buckberry-Chamberlain methods, significant differences in median between Spitalfields and the South African collections were found in the 50 to 59 year group and over, but no other differences were consistent. The results for each age determination method and age group are in Appendix 6.

The Buckberry-Chamberlain scored traits were also divided into ten-year age groups for analysis; tables of results are in Appendix 7. The most interesting result for the scored traits were significant differences in median for surface texture between Spitalfields and the South African collections, again from the 50 to 59 year group and older.

4.12 Mean Age per Phase/Score

One-way ANOVAs were used to detect differences in mean age per phase between collections for each of the ageing methods, for the sexes pooled and for females and males separately. ANOVA, however, is sensitive to unequal population variances unless the sample sizes are equal and, in this case, there is an increased possibility of giving a significant difference in means when none exists (Type I error). Accordingly, Levene's test was used to look for equality of variance, where the null hypothesis is that population variances are equal; if the p-value is 0.05 or less, the null hypothesis is rejected, and population variances are not equal. Where there is inequality of variance, and sample sizes are not approximately equal, if smaller samples have the larger variances, than there is an increased possibility of ANOVA returning significant differences that do not actually exist; similarly, if the smaller samples have the smaller population variances, and larger samples have larger population variances, than ANOVA is less likely to return a significant difference that does exist. Where variance is not equal, the Robust Tests of Equality of Means (hereafter called RTEM) can be used to determine whether significant differences exist between means instead of ANOVA. Where significant differences in mean were found, Tukey's HSD (a post-hoc test) indicates the location of differences. For some phases or scores, a second ANOVA was performed to exclude collections with only one individual in that phase/score.

4.12.1 Suchey-Brooks Mean Age by Phase

The first ageing method examined was the Suchey-Brooks pubic symphysis method, for the sexes pooled (results are presented in Table 4.90, below). Results for the female-only data and male-only data (Tables 4.91 and 4.92) are presented after that of the sexes pooled.

	Phase					
	1	2	3	4	5	6
Levene's test	**.007**	**.000**	.464	**.010**	.124	.684
ANOVA	.216	.007	.688*/.638**	.924	.098	**.024**
RTEM	.458	.091	--	.886	--	--

Table 4.90. Suchey-Brooks phases, sexes pooled: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case

Unequal variances were found for phases 1, 2 and 4. For these phases, RTEM was the appropriate measure of significance for the difference of means due to variation in sample size. For phase 3, the Grant Collection had only one individual, so a second ANOVA was run excluding the Grant individual. A significant difference in mean was found for phase 6 (see Figure 4.21); post-hoc tests indicate that the significant difference was between Dart and Grant.

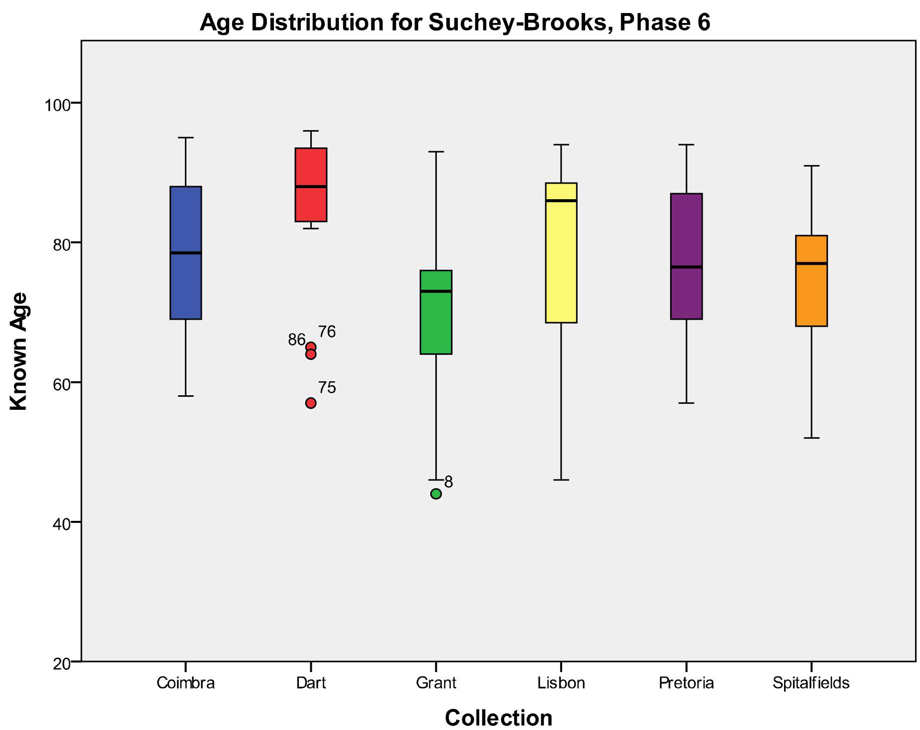


Figure 4.21. Boxplot of known age distribution for Suchey-Brooks phase 6 for the sexes pooled, by collection

For females, second ANOVAs for phases 1 (excluding Spitalfields) and 2 (excluding Grant) were run. Unequal variance was found for phase 2; this was also the only phase with a significant difference in mean. The significant difference seemed to be between Lisbon and Spitalfields.

	Phase					
	1	2	3	4	5	6
Levene's test	.098	**.000**	.108	.344	.756	.831
ANOVA	.618*/.532**	**.048*/.042****	.943	.885	.315	.398
RTEM	--	**.003**	--	--	--	--

Table 4.91. Suchey-Brooks phases, females: ANOVA and test of variance
RTEM: included when necessary;*including all collections; **excluding those with only one case

For males, exclusions from second ANOVAs were Lisbon and Pretoria (phase 1), Coimbra and Spitalfields (phase 2), and Grant (phase 3). Unequal variance and small sample sizes occurred in phase 1, so RTEM was used to measure difference of mean. When the appropriate measures of differences of mean were observed, no significant differences were found for males only.

	Phase					
	1	2	3	4	5	6
Levene's test	**.031**	.122	.439	.059	.120	.085
ANOVA	.538*/.300**	**.020***/.228**	.265*/.220**	.914	.433	.134
RTEM	.458	--	--	--	--	--

Table 4.92. Suchey-Brooks phases, males: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case

4.12.2 Meindl-Lovejoy Mean Age by Phase

The Meindl-Lovejoy auricular surface method was next analysed for differences in mean age per phase. The results for the sexes pooled are presented in Table 4.93. For the 20-24 year phase, a second ANOVA excluded Spitalfields and Grant. Second ANOVAs for the 25-29 year phase (excluding Grant) and the 60+ phase (excluding Pretoria) were also performed. Unequal variances and small or varied sample sizes for the 25-29, 30-34, 35-39 and 50-60 year phases required RTEMs to be used for testing difference of means.

Significant differences in means were found in the 30-34, 35-39, 40-44 and 50-60 year phases (see Figures 4.22 to 4.25; for all boxplots, the middle bar shows the median, the box shows the interquartile range, and the whisker shows the range). For the 30-34 and 35-39 year phases, the differences were between Dart and Grant. The 35-39 year phase also had significant differences between Dart and Spitalfields and Pretoria and Grant. The 40-44 year phase differences were between Spitalfields and Lisbon, while the 50-60 year phase differences were between Grant and Coimbra.

	Phase							
	20-24	25-29	30-34	35-39	40-44	45-49	50-60	60+
Levene's test	.083	**.003**	**.014**	**.001**	.423	.350	**.000**	.164
ANOVA	.338*/.198**	.232*/.230**	.284	**.000**	**.008**	.403	**.042**	.325*/.238**
RTEM	--	.436	**.005**	**.000**	--	--	**.020**	--

Table 4.93. Meindl-Lovejoy phases, sexes pooled: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case

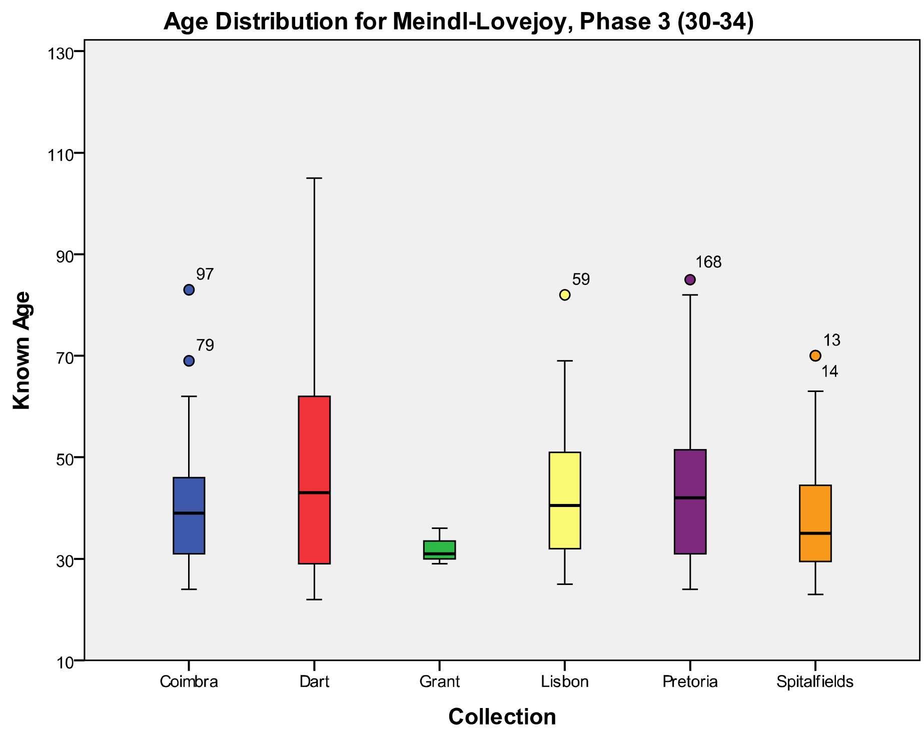


Figure 4.22. Boxplot of known age distribution for Meindl-Lovejoy 30-34 year phase for the sexes pooled, by collection

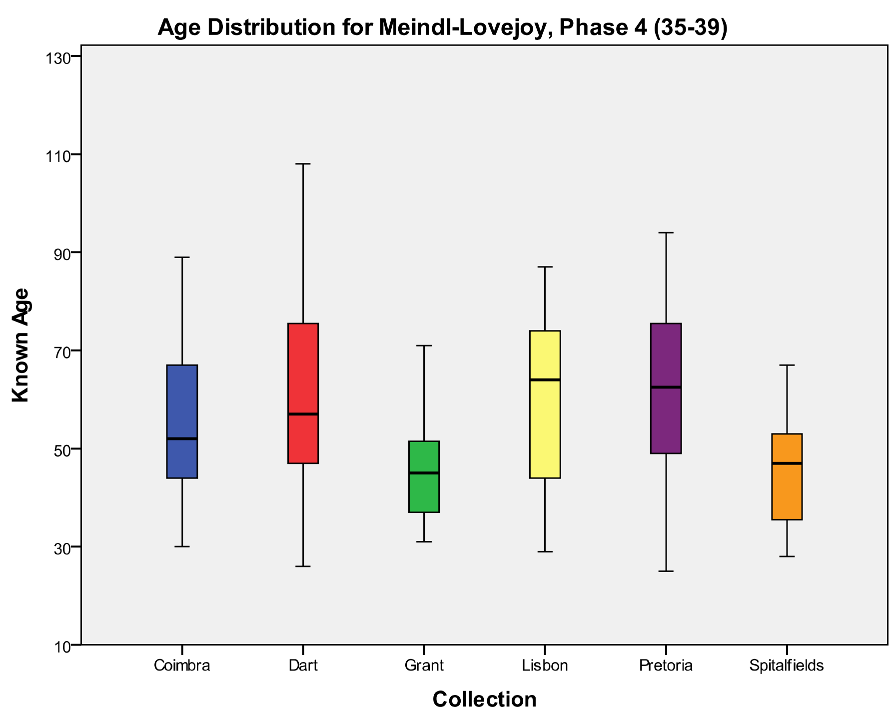


Figure 4.23. Boxplot of known age distribution for Meindl-Lovejoy 35-39 year phase for the sexes pooled, by collection

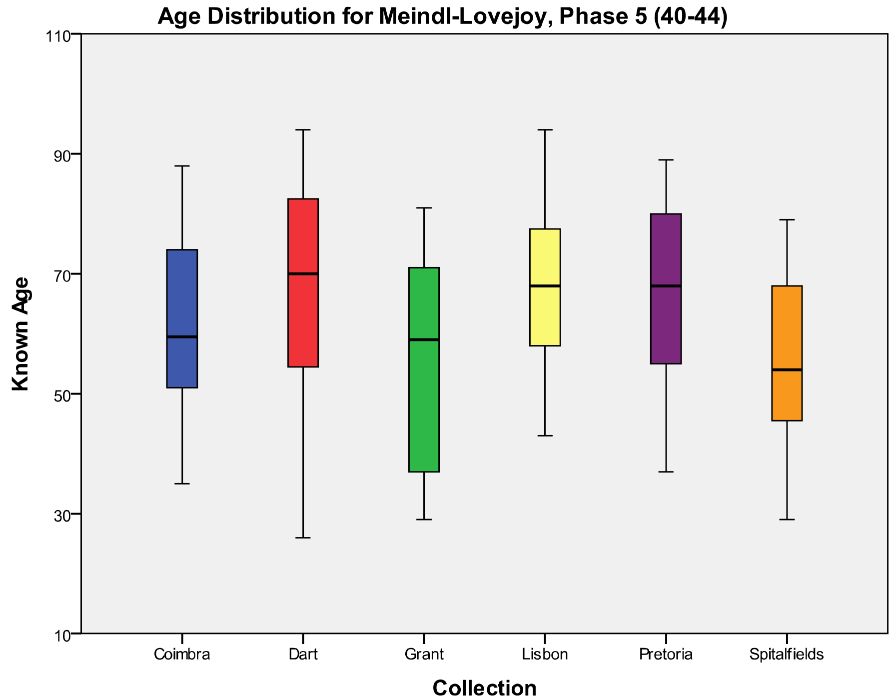


Figure 4.24. Boxplot of known age distribution for Meindl-Lovejoy 40-44 year phase for the sexes pooled, by collection

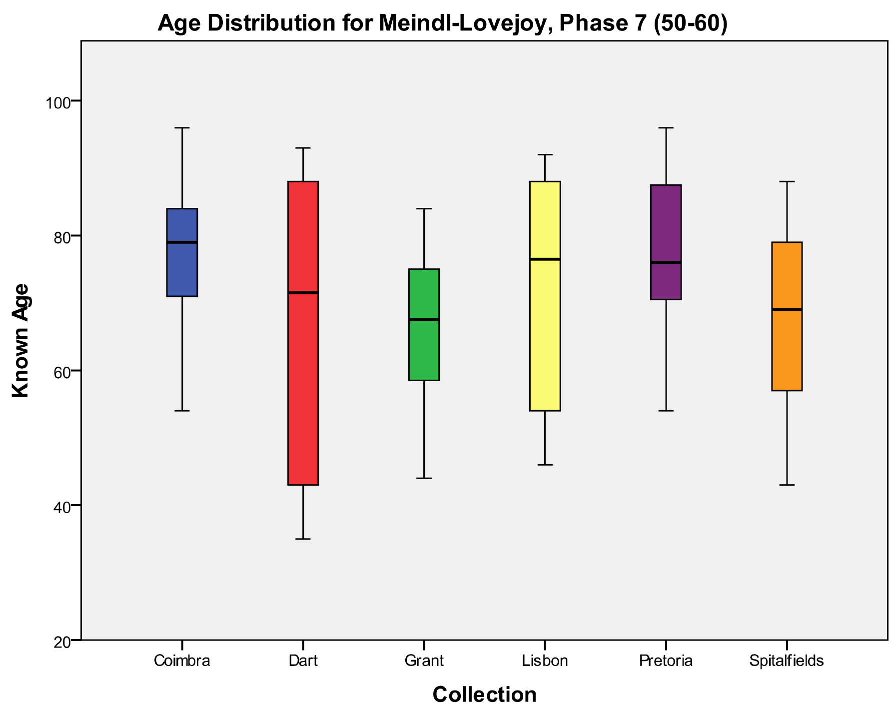


Figure 4.25. Boxplot of known age distribution for Meindl-Lovejoy 50-60 year phase for the sexes pooled, by collection

For the female data (Table 4.94), second ANOVAs for the 20-24 year phase (excluding Coimbra, Spitalfields and Grant), for the 25-29 and 40-44 year phases (excluding Grant), and for the 60+ phase (excluding Pretoria) were performed. Unequal variance and varied sample sizes for the 25-29 and 30-34 year phases required the use of RTEM instead of ANOVA. Only the 40-44 year phase had a significant difference in means; post hoc tests indicated that the differences were between Spitalfields compared to Lisbon and Pretoria.

	Phase							
	20-24	25-29	30-34	35-39	40-44	45-49	50-60	60+
Levene's test	.503	**.003**	**.049**	.662	.369	.319	.098	.745
ANOVA	.805*/ .625**	.408*/ .439**	.302	.054	**.012*/ .028****	.244	.081	.708*/ .561**
RTEM	--	.674	.407	--	--	--	--	--

Table 4.94. Meindl-Lovejoy, females only: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case

For the male data (Table 4.95), second ANOVAs for the 20-24 year phase (excluding Coimbra) and the 60+ phase (excluding Dart) were performed. Unequal variance and varied sample sizes for the 35-39 year phase required the use of RTEM. Significant differences in mean were found in the 35-39 (between both Grant and Spitalfields compared to Dart and Pretoria), and 50-60 year phases (between Dart compared to Coimbra and Pretoria).

	Phase							
	20-24	25-29	30-34	35-39	40-44	45-49	50-60	60+
Levene's test	.165	.399	.327	**.000**	.797	.262	.843	.069
ANOVA	.502*/ .449**	.608	.828	**.000**	.175	.609	**.036**	.380*/ .525**
RTEM	--	--	--	**.000**	--	--	--	--

Table 4.95. Meindl-Lovejoy phases, males only: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case

4.12.3 Buckberry-Chamberlain Mean Age by Phase

The Buckberry-Chamberlain auricular surface phases were tested next for differences in mean age per phase. The results for the sexes pooled are presented in Table 4.96. A second ANOVA was done for phase III (excluding Grant). Unequal variance and varying sample sizes for phase III and V were found; RTEM was used for those phases instead of ANOVA.

Significant differences in mean were found between phases III, IV and V (see Figures 4.26 to 4.28). The phase III difference was between Pretoria and Coimbra, and the phase IV difference was between Spitalfields and Dart. Phase V differences were between Spitalfields compared to Dart, Lisbon and Pretoria, and between Grant compared to Lisbon and Pretoria.

	Phase					
	II	III	IV	V	VI	VII
Levene's test	.687	**.003**	.429	**.002**	.353	.554
ANOVA	.246	.060*/.075**	**.027**	**.000**	.216	.113
RTEM	--	**.039**	--	**.000**	--	--

Table 4.96. Buckberry-Chamberlain phases, sexes pooled: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case

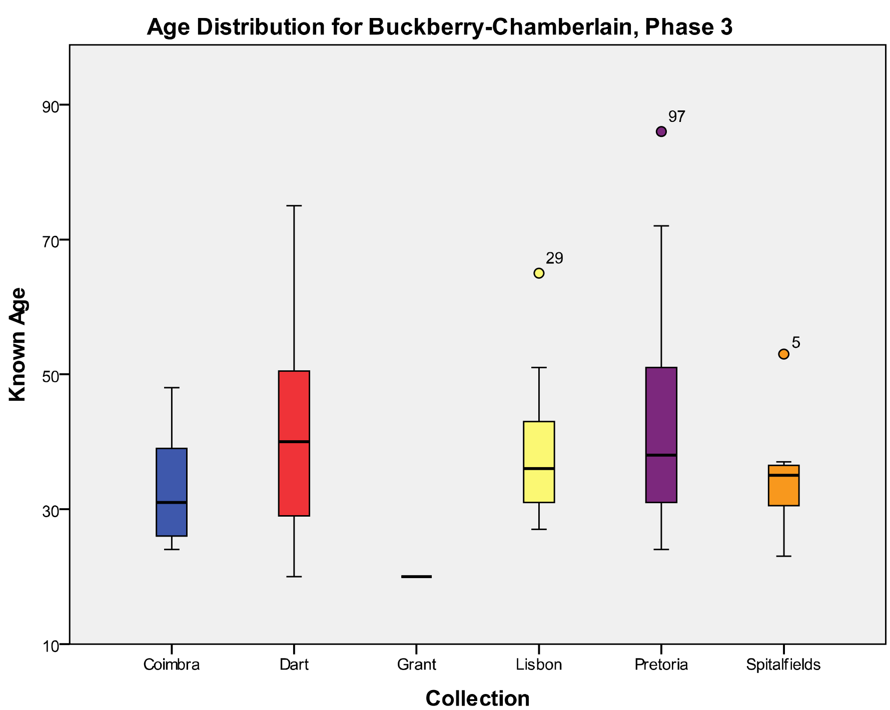


Figure 4.26. Boxplot of known age distribution for Buckberry-Chamberlain phase III for the sexes pooled, by collection

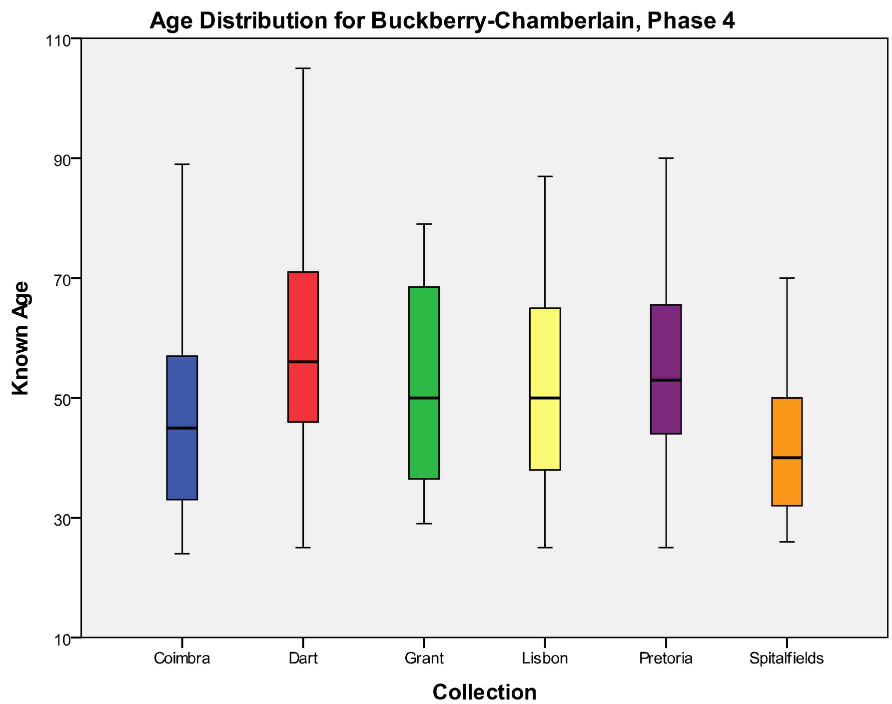


Figure 4.27. Boxplot of known age distribution for Buckberry-Chamberlain phase IV for the sexes pooled, by collection

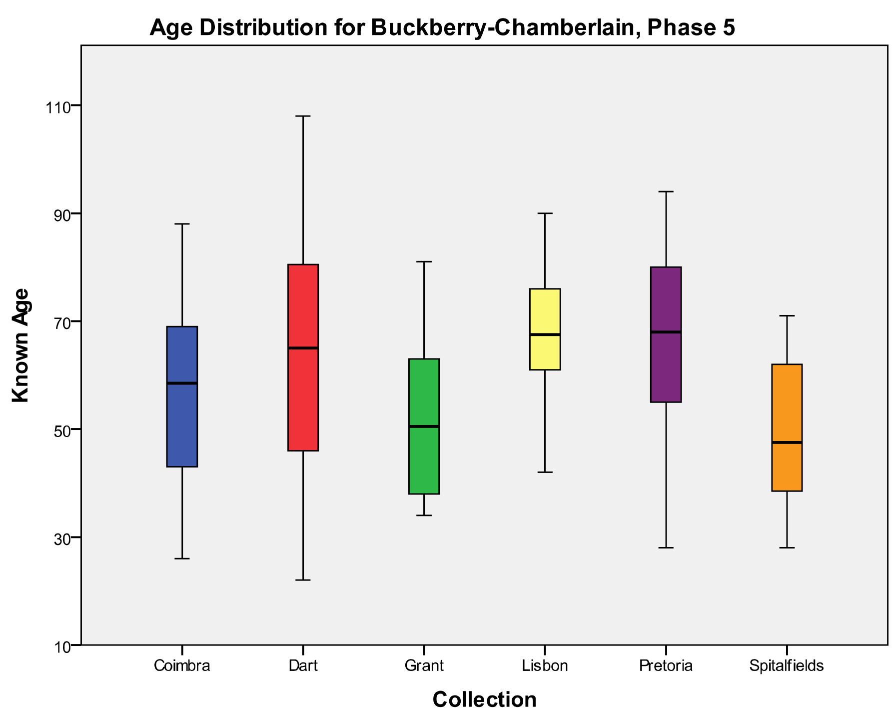


Figure 4.28. Boxplot of known age distribution for Buckberry-Chamberlain phase V for the sexes pooled, by collection

For females (Table 4.97), RTEM was used for phase III, due to unequal variance and varying sample sizes. Phases III and V had significant differences in means; the phase III difference was between Coimbra and Pretoria, while the phase V difference was between Pretoria and Spitalfields.

	Phase					
	II	III	IV	V	VI	VII
Levene's test	.506	**.002**	.782	.126	.485	.074
ANOVA	.504	.074*/.090**	.230	**.035**	.389	.283
RTEM	--	**.029**	--	--	--	--

Table 4.97. Buckberry-Chamberlain phases, females only: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case

In the males-only results (Table 4.98), unequal variance and varying sample sizes were found for phases II and V; RTEM was thus used instead of ANOVA for these phases. Only phase V has a significant difference in mean. The significant differences were between Lisbon compared to Grant and Spitalfields.

	Phase					
	II	III	IV	V	VI	VII
Levene's test	**.021**	.348	.470	**.024**	.295	.638
ANOVA	.062	.694	.096	**.005**	.512	.327
RTEM	.058	--	--	**.000**	--	--

Table 4.98. Buckberry-Chamberlain phases, males only: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case

4.12.4 Lateral-Anterior Cranial Suture Closure Mean Age by Phase

The results for the sexes pooled for lateral-anterior suture closure are presented first (Table 4.99). A second ANOVA (excluding Grant) was done for phase 3. For phase 4, a t-test to test the means of Coimbra and Lisbon was done as these were the only two collections with more than one individual in this phase. RTEM was necessary only for phase 8, with unequal variance and small sample sizes. A significant difference in mean was found in phase 6 (Figure 4.29); post hoc tests showed that the lowest p-values were in Grant compared to Coimbra and Pretoria.

	Phase					
	3	4	5	6	7	8
Levene's test	.135	.082	.590	.689	.089	**.001**
ANOVA	.420*/.263**	.795*/.529***	.996	**.016**	.757	.875
RTEM	--	--	--	--	--	.951

Table 4.99. Lateral-anterior suture phases, sexes pooled: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case; ***t-test, only two collections

The females-only results are presented in Table A8.1, Appendix 8. For phases 3 and 4, t-tests were used instead of ANOVA between Coimbra and Lisbon, as these were the only two collections with more than one individual. Insufficient females were in phase 8 for analysis of difference of means. No significant differences for mean were found for the females alone.

In the male data (Table 4.100), unequal variances and small sample sizes were found in phases 4 and 8; a t-test was instead used for phase 4 (between Coimbra and Lisbon), and RTEM was used for phase 8. Only phase 6 had a significant difference in mean; the significant difference was between Dart and Coimbra.

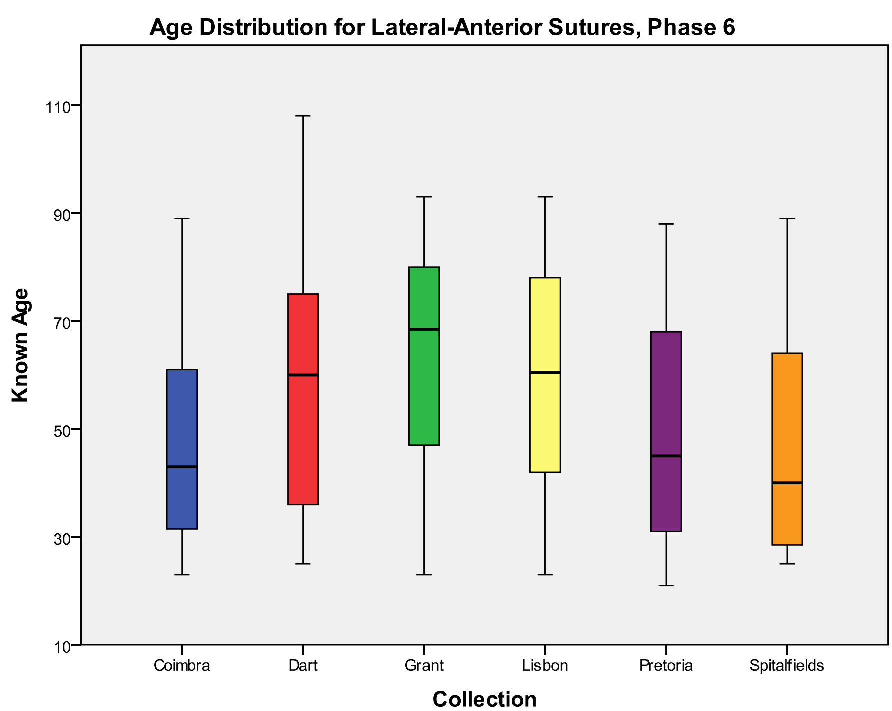


Figure 4.29. Boxplot of known age distribution for lateral-anterior suture phase 6 for the sexes pooled, by collection

	Phase					
	3	4	5	6	7	8
Levene's test	--	**.000**	.123	.068	.088	**.001**
ANOVA	--	.541*/.451***	.564	**.010**	.950	.875
RTEM	--	.492***	--	--	--	.951

Table 4.100. Lateral-anterior sutures, males only: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case; ***t-test, only two collections; -- : n/a

4.12.5 Vault Cranial Suture Closure Mean Age by Phase

Few significant differences were found in mean for the vault suture phases; as such, p-values for the statistical tests are in Tables A8.2 to A8.4, Appendix 8.

For the sexes pooled, a second ANOVA was done for phase 4 (excluding Grant). Unequal variance and varied sample sizes for phase 5 required the use of RTEM instead of ANOVA. No significant differences in mean were found for any phases.

For females only, a second ANOVA for phase 5 excluded Grant. Only phase 6 had a significant difference in mean for the female data. Post hoc tests indicated the significant differences were in Dart compared to Coimbra and Pretoria.

For males only, a second ANOVA was done for phases 3 (excluding Spitalfields) and 4 (excluding Grant). RTEM was used for phases 3 and 5 due to unequal variances and varied sample sizes. No significant differences were found in means for the male data.

4.12.6 Sternal End of Fourth Rib Mean Age by Phase

Relatively few individuals had fourth ribs with intact sternal ends, so ANOVAs could not be run for every phase; as such, even when there is a "*" next to a p-value, it actually refers to all collections with any individuals in that phase, not necessarily all six collections. As few significant differences in mean were found; tables of resulting p-values are in Tables A8.5 to A8.7, Appendix 8.

For the sexes pooled, second ANOVAS were done for phases 9 and 10 (both excluding Grant), and a t-test was done for phase 13 between Coimbra and Dart, as the other collections had insufficient numbers of individuals. RTEM was required for phases 5, 6 and 11 instead of ANOVAs due to unequal variance and varied sample sizes.

For females, second ANOVAs were done for phases 5 (excluding Pretoria) and 10 (excluding Dart); a t-test was done for phase 11 between Dart and Spitalfields. RTEM was used for phases 6 and 7 due to unequal variances and varied sample sizes. The only significant difference was found in phase 10; the difference in means was between Lisbon and Pretoria.

For males only, a t-test was done for phase 6 (between Spitalfields and Pretoria), and second ANOVAs were done for phases 9 and 10 (both excluding Grant). Another t-test was done for phase 13, between Coimbra and Dart. Unequal variance and small sample sizes were found for phase 6; RTEM was used here instead of ANOVA. Only phase 10 had a significant difference in mean for the male data. The differences seem to be between Pretoria and Coimbra, and Coimbra and Spitalfields (although the post-hoc test p-values were not significant).

4.13 Buckberry-Chamberlain Trait Scores

4.13.1 Transverse Organisation Mean Age by Score

The individual skeletal features scored for the Buckberry-Chamberlain method were also subjected to ANOVAs and tests of equality of variance for mean ages. The scores for transverse organisation were examined first. As only one significant difference in mean was found, the tables of p-values are presented in Tables A8.8 to A8.10, Appendix 8.

For the sexes pooled, RTEM was used for scores of 2 due to unequal variance and varied sample sizes. No significant differences in mean age were found for transverse organisation scores for the sexes pooled. For females only, RTEM was again used for scores of 2 due to unequal variance and varied sample sizes. No significant differences in mean were found. For males only, the only significant difference in mean was for scores of 5. Post-hoc tests indicated that the difference was in Spitalfields compared to Dart and Pretoria.

4.13.2 Surface Texture Mean Age by Score

The surface texture scores were analysed next for differences in mean; tables (A8.11 to A8.13) with p-values are given in Appendix 8.

For the sexes pooled, RTEM was used for testing means of scores of 2, 3 and 4, due to unequal variances and varied sample sizes. Significant differences were found for scores of 3 and 4. The differences for scores of 3 were between Dart and Spitalfields, while differences for scores of 4 were in Lisbon compared to Grant and Spitalfields. No significant differences were found in the females-only means. For males only, RTEM was used for scores of 2 and 3 due to unequal variance and varied sample sizes. A significant difference in mean was found for scores of 3 only; the lowest p-value (non-significant, at 0.084) given by post-hoc tests was for Spitalfields and Dart.

4.13.3 Microporosity Mean Age by Score

Microporosity scores were analysed next for differences in means; Tables A8.14 to A8.16 in Appendix 8 provide all p-values.

For the sexes pooled, RTEM was used for scores of 1 and 2, as variances were unequal and sample sizes varied widely. Significant differences in mean were found for all possible scores. The differences in scores of 1 were between Grant and Lisbon; differences in scores of 2 in Dart compared to Lisbon and Coimbra; differences in scores of 3 were between Spitalfields and Lisbon. No significant differences were found in the females-only data. For males only, RTEM was used for scores of 1 and 2 due to unequal variances and widely varying sample sizes. Significant differences were found in mean for scores of 1 and 2. The lowest p-value (non-significant, 0.099) given by post-hoc tests for scores of 1 was for Pretoria and Lisbon. The differences for scores of 2 were in Dart compared to Coimbra and Lisbon.

4.13.4 Macroporosity Mean Age by Score

The macroporosity scores for the sexes pooled were analysed next (all p-values are in Tables A8.17 to A8.19, Appendix 8). For the sexes pooled, unequal variance and varied sample sizes for scores of 1 required the use of RTEM instead of ANOVA. Significant differences in mean were found for scores of 2; the difference seems to be between Pretoria and Coimbra (the lowest p-value found by post-hoc tests, at 0.055). For females alone, no significant differences were found in variance or mean. For males, RTEM was used for scores of 1 due to unequal variance and widely varying sample sizes. No significant differences in mean were found.

4.13.5 Apical Change Mean Age by Score

The apical change scores for the sexes pooled were analysed next (for all p-values, see Tables A8.20 to A8.22, Appendix 8). A significant difference in means was detected in scores of 2. Post hoc testing indicated that the difference was between Pretoria and Spitalfields. No significant differences were found in mean for females or males only.

4.14 Overall vs Subjective Age Estimates

The results for both summary age estimates were compared to determine which of the two performed better – the overall estimates (using only formal ageing methods) or subjective age estimates (incorporating informal age indicators alongside formal ageing methods). A correct age estimate is where the estimated age range includes the known age. If the estimate was only one year from the known age, it was still considered incorrect. Generally, the overall age estimates did not perform as well as the subjective age estimates. The Grant Collection (the first collection visited) was the only sample for which the subjective estimates did not improve the percentage of correct age estimates; for all other collections, the improvement of subjective over overall estimates was around 20%. The Grant Collection was visited again, to test 20 individuals (just over 20% of the original Grant sample) for intraobserver error. For the second visit, subjective age estimation improved on the overall estimates in terms of percentage of correct age estimates.

Subjective estimates from the Pretoria Collection sample had the highest improvement compared to overall estimates, from 39.9% correct (overall estimates) to 68.2% correct (subjective estimates) – a 28.3% difference. For overall estimates, the highest percentage of correctly estimated ages was for the Coimbra sample, at 49.3% correct, and the lowest was for the Dart Collection, at 54.7% correct. The highest percentage of correctly estimated subjective ages was for the Lisbon sample, at 70.6% correct; the lowest was for the second Grant visit, at 50.0%, but sample size was only 20 individuals for the intraobserver error test. The lowest proportion of correctly estimated ages using the subjective method was higher than the highest proportion of correctly estimated ages using the overall method. Table 4.101 below shows percentages of correct estimates and the difference in percentage between the two summary methods.

	Correct Age Estimate (%)		Difference (%)
	Overall	Subjective	
Coimbra	49.3	67.1	+ 17.8
Dart	31.4	54.7	+23.3
Grant	38.6	28.9	-9.7
Lisbon	48.0	70.6	+22.6
Pretoria	39.9	68.2	+28.3
Spitalfields	42.5	64.2	+21.7
Grant 2	35.0	50.0	+15.0

Table 4.101. Percentages of correct age estimates using overall and subjective methods, and difference between them. [For difference (%), a plus sign indicates an improvement in age estimates from overall to subjective; a negative sign indicates correct proportions of age estimates decreased from overall to subjective.]

To analyse whether these differences in correct age estimates were significant, McNemar's chi square tests were done – the overall and subjective estimates for each individual were treated as paired observations for this test. This also gives a cross-tabulation of whether paired estimates were both correct, both incorrect, were incorrect the first time (i.e. overall estimate) and correct the second time (i.e. subjective estimate) or vice versa (Table 4.102).

In all cases except both Grant Collection visits, there were significant differences between the totals of correctly estimated ages using the overall compared to subjective estimates. That is, the McNemar tests show that the subjective method significantly improved the number of correct age estimates compared to the overall method (except for Grant).

	Both Correct		Both Incorrect		Correct to Incorrect		Incorrect to Correct		McNemar p-value
	n	%	n	%	n	%	n	%	
Coimbra	56	40.0	33	23.6	13	9.3	38	27.1	**.001**
Dart	35	22.0	57	35.8	15	9.4	52	32.7	**.000**
Grant	15	18.1	42	50.6	17	20.5	9	10.8	.169
Lisbon	64	43.8	37	25.3	6	4.1	39	26.7	**.000**
Pretoria	51	34.4	39	26.4	8	5.4	50	33.8	**.000**
Spitalfields	47	37.0	37	29.1	7	5.5	36	28.3	**.000**
Grant 2	3	15.0	6	30.0	4	20.0	7	35.0	.549

Table 4.102. Percentages of age estimates that remained the same or changed from overall to subjective estimates and significance of differences

The individuals who were incorrectly aged with both methods did not mark a decrease in reliability from one method to another; proportions range from 23.6% for Coimbra to a high of 50.6% for Grant. After the Grant Collection, the Dart Collection had the lowest proportion of individuals aged correctly with both the overall and subjective methods – there was agreement (and correct ages) for only 22.0% of individuals. The South African collections showed the most improvement (individuals with incorrect estimates using the overall method but correct estimates using the subjective method), bar the second Grant visit – at 32.7% for Dart and 33.8% for Pretoria. It is not surprising that the second Grant visit yielded the most improvement, because the overall method did not perform well for either Grant visit and there was a high proportion of incorrect estimates for the first visit. The proportions of individuals aged correctly with the overall method but incorrectly with the subjective method were all quite low (except for Grant), ranging from 4.1% (for Lisbon) to 9.4% (for Dart).

Although the subjective method improved age estimates in general, it was helpful to consider the effect of the negative results (the individuals who were correctly aged using the overall method, but incorrectly using the subjective method) in examining the improvement of the subjective method over the overall method. To take these negative results into account, a percentage of change was calculated, where positive percentages reflected improvement and negative percentages reflected worsening. The number of individuals in the "Correct to Incorrect" category were subtracted from the number of individuals in the "Incorrect to Correct" category and divided by the number of individuals per sample to calculate a percentage of total change (Table 4.103).

	(I to C) – (C to I)	% change
Grant (1)	-8	-9.6
Spitalfields (2)	29	22.8
Coimbra (3)	25	17.8
Lisbon (4)	33	22.6
Dart (5)	37	23.3
Pretoria (6)	42	28.4
Grant 2 (7)	3	15

Table 4.103. Difference between overall to subjective method – proportion of correct estimates, accounting for negative change.
(I to C): the number of individuals incorrectly aged by overall method, but correctly aged by subjective method; (C to I): the number of individuals correctly aged by overall method, but incorrectly aged by subjective method; numbers in brackets beside collections indicate the order in which data were collected.

When the order of data collection was examined, there was a general trend of an increase in percentage of change, with the exception of the second Grant visit. However, the small intraobserver error sample size (*n* = 20) affected these changes. Otherwise, from the first Grant visit to the Pretoria visit, the overall trend was one of improvement.

Results were also divided by age group; the number of individuals, and the number and percentage of correct estimates were examined by known ages, grouped into decades (as sampling was based on decade of age-at-death). As no particular trends or differences were observed between sexes of the same collection in terms of percentages of correct estimates, the details of the female- and male-only results are presented in Appendix 9. Inaccuracy, standard deviation, and bias of age estimates were also calculated for overall and subjective estimates for each collection and grouped into decade of age-at-death. These measures are now standard in examining efficacy of age determination methods, to look for age-related patterns in age estimation error.

4.14.1 Grant Collection

The Grant Collection overall estimates were low throughout the age groups (see Table 4.104) and the lowest percentages of correct estimates were in the oldest age groups (from the 70 to 79 age group and older). There was also a decrease in correct subjective estimates (Table 4.105) in the older age groups from the 60 to 69 group onwards.

Inaccuracy, standard deviation and bias were next examined for the sexes pooled for the Grant Collection (Table 4.106, Figures 4.30 and 4.31). For all three of these measures, for both the overall and subjective estimates, there was an increase with age, beginning from the 70 to 79 year group. Inaccuracy, standard deviation and bias were all greater for the subjective estimates

Known Age Group	Number of Individuals	Correct Age Estimates (*n*)	Correct Age Estimates (%)
20-29	5	2	40.0
30-39	11	6	54.6
40-49	11	3	27.3
50-59	--	--	--
60-69	15	8	53.3
70-79	18	3	16.7
80-89	10	0	0.0
90-99	--	--	--
100+	--	--	--
Total	83	32	38.6

Table 4.104. Grant Collection, numbers and percentages of correct overall estimates

Known Age Group	Number of Individuals	Correct Age Estimates (*n*)	Correct Age Estimates (%)
20-29	5	0	0.0
30-39	11	9	81.8
40-49	11	5	45.4
50-59	--	--	--
60-69	15	2	13.3
70-79	18	2	11.1
80-89	10	1	10.0
90-99	--	--	--
100+	--	--	--
Total	83	24	28.9

Table 4.105. Grant Collection, numbers and percentages of correct subjective estimates

compared to the overall estimates for the oldest age group (the most difficult to age). Bias tended to be positive (overageing) for the younger age groups, and negative (underageing) for the older age groups using both overall and subjective estimates, from the 60 to 69 and 40 to 49 year age groups, respectively.

When total inaccuracy and standard deviation were considered, both were slightly greater for the subjective estimates compared to the overall estimates. Total bias was absolutely smaller for the subjective estimates compared to that of the overall estimates, but subjective estimates tended to underage, while overall estimates tended to overage on average. The youngest age groups (for both overall and subjective estimates) had the greatest amount of overageing, which then decreased by age group until the point (age group) at which underageing begins. Underageing then increased with age group, forming a parabolic pattern. There was a small fluctuation in the overall estimates (for the 30 to 39 and the 40 to 49 age groups), and the fact that the next age group (50 to 59) did not have enough individuals to calculate bias does not help in understanding whether this was a random fluctuation or a break in the pattern.

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	5.40	4.20	6.99	0.84	5.40	4.20
30-39	2.55	1.18	3.24	2.86	2.55	1.18
40-49	4.36	3.73	4.78	3.88	3.82	-1.73
50-59	n/a	n/a	n/a	n/a	n/a	n/a
60-69	3.80	6.33	4.93	4.43	-3.80	-4.60
70-79	9.11	8.50	7.53	6.19	-9.11	-7.83
80-89	15.40	18.40	7.96	10.35	-15.40	-18.40
90-99	n/a	n/a	n/a	n/a	n/a	n/a
100+	n/a	n/a	n/a	n/a	n/a	n/a
Total (mean)	6.13	6.93	7.11	7.47	7.11	-5.39

Table 4.106. Grant Collection, inaccuracy, standard deviation and bias for each age group, sexes pooled

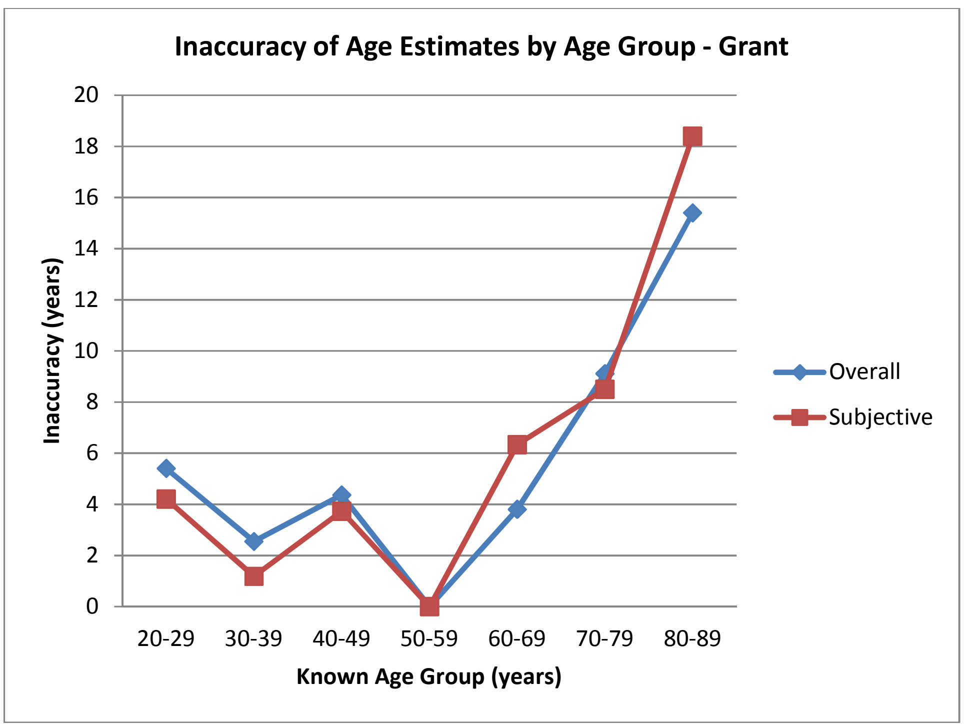


Figure 4.30. Inaccuracy of age estimates by age group for the Grant Collection

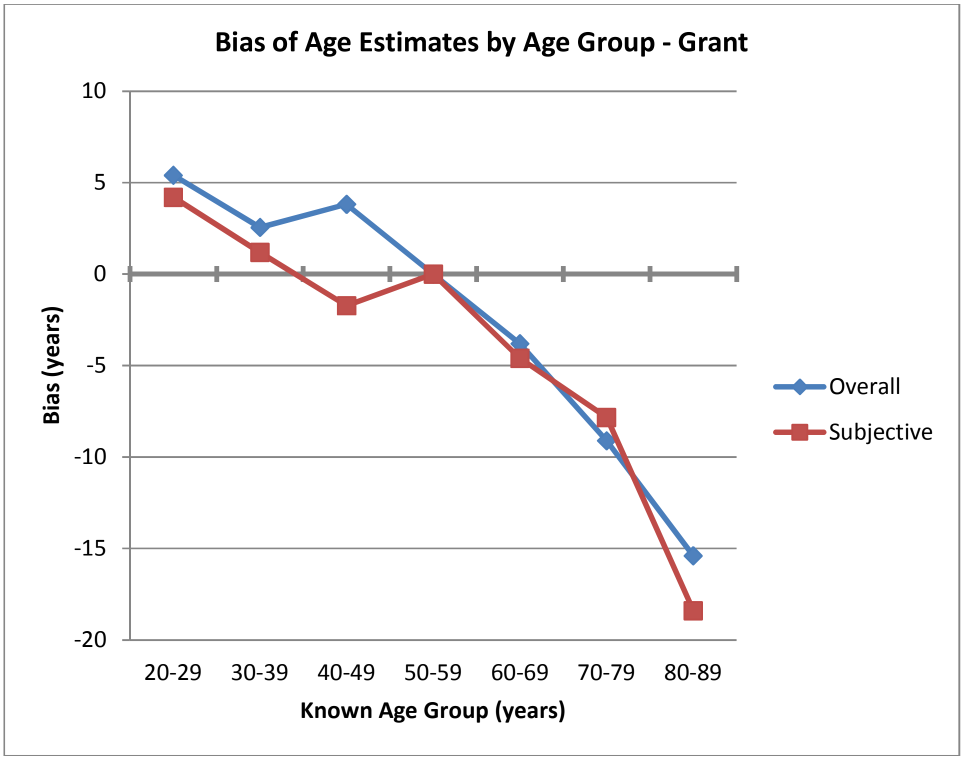


Figure 4.31. Bias of age estimates by age group for the Grant Collection

4.14.2 Spitalfields Collection

The pattern of correct overall age estimates by age group mirrored that of bias (the parabolic pattern), with a low initial percentage of correct estimates for the youngest age group, which increased steadily (here, the highest percentage for overall estimates was the 40 to 49 year age group) until a reversal in the middle to older age groups, decreasing in percentage after that. The decrease began for the overall estimates at the 50 to 59 year age group. No individuals aged 70 year and older were aged correctly. Full details are in Table 4.107, below.

Known Age Group	Number of Individuals	Correct Age Estimates (*n*)	Correct Age Estimates (%)
20-29	14	5	35.7
30-39	21	15	71.4
40-49	20	17	85.0
50-59	19	11	57.9
60-69	21	6	28.6
70-79	18	0	0.0
80-89	13	0	0.0
90-99	--	--	--
100+			
Total	127	54	42.5

Table 4.107. Spitalfields Collection, numbers and percentages of correct overall estimates

The subjective Spitalfields estimates followed the same pattern of increasing percentages of correct estimates until a reversal after the 30 to 39 year age group. Percentages then decreased with age until, interestingly, the 80 to 89 year age group, which had higher percentages of correct estimates than the two preceding age groups. Details are in Table 4.108, below.

For the overall estimates for the sexes pooled, inaccuracy, standard deviation and bias all followed the aforementioned parabolic pattern, first decreasing steadily, then increasing steadily from the 50 to 59 age group (see Table 4.109 and Figures 4.32 and 4.33). For bias, overageing again occurred in the younger age groups before switching to underageing in the 60 to 69 age group, which increased with age. There was a minor fluctuation in the 40 to 49 and 50 to 59 year groups.

The subjective age estimate values for Spitalfields did not follow this pattern. Inaccuracy fluctuated over the entire age range; inaccuracy was lowest for the 30 to 39 group and highest for the 70 to 79 age group. The standard deviations for the subjective estimates also fluctuated across the age groups. Bias for the subjective age estimates increased steadily from the 20 to 29 age group to the 70 to 79 age group, decreasing for the 80 to 89 age group. The switch from positive bias to negative bias was still present, and occurred in the 60 to 69 year age group.

Known Age Group	Number of Individuals	Correct Age Estimates (*n*)	Correct Age Estimates (%)
20-29	15	10	66.7
30-39	21	17	81.0
40-49	20	5	75.0
50-59	20	12	60.0
60-69	21	11	52.4
70-79	21	10	47.6
80-89	15	11	73.3
90-99	--	--	--
100+	--	--	--
Total	134	86	64.2

Table 4.108. Spitalfields Collection, numbers and percentages of correct subjective estimates

The greatest amount of bias was -15.54, for the 80 to 89 year olds, using the overall method; the subjective value for the same age group was -1.93. Bias and inaccuracy were lower for the subjective estimates than they were for the overall estimates in nearly every age group (except the 40 to 49 and 50 to 59 year age groups). The total mean values for inaccuracy, standard deviation and bias were all lower for the subjective estimates compared to the overall estimates.

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	7.86	0.93	8.36	1.91	7.86	0.27
30-39	3.00	0.48	5.83	1.25	3.00	0.48
40-49	0.70	1.65	1.95	3.92	-0.40	1.05
50-59	2.00	3.30	2.96	4.92	0.42	1.40
60-69	4.67	2.29	4.84	3.07	-4.48	-1.81
70-79	8.22	4.10	5.81	8.40	-8.22	-3.52
80-89	15.54	1.93	7.39	4.03	-15.54	-1.93
90-99	--	--	--	--	--	--
100+	--	--	--	--	--	--
Total (mean)	5.46	2.25	6.97	4.78	-2.30	-0.70

Table 4.109. Spitalfields Collection, inaccuracy, standard deviation and bias for each age group, sexes pooled

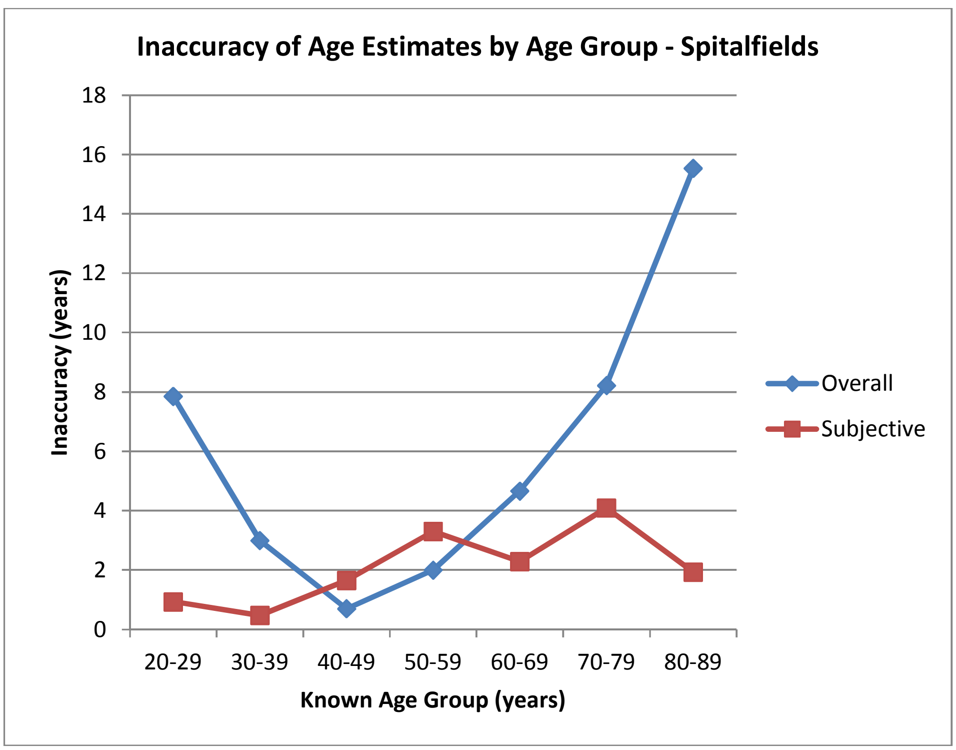


Figure 4.32. Inaccuracy of age estimates by age group for the Spitalfields Collection

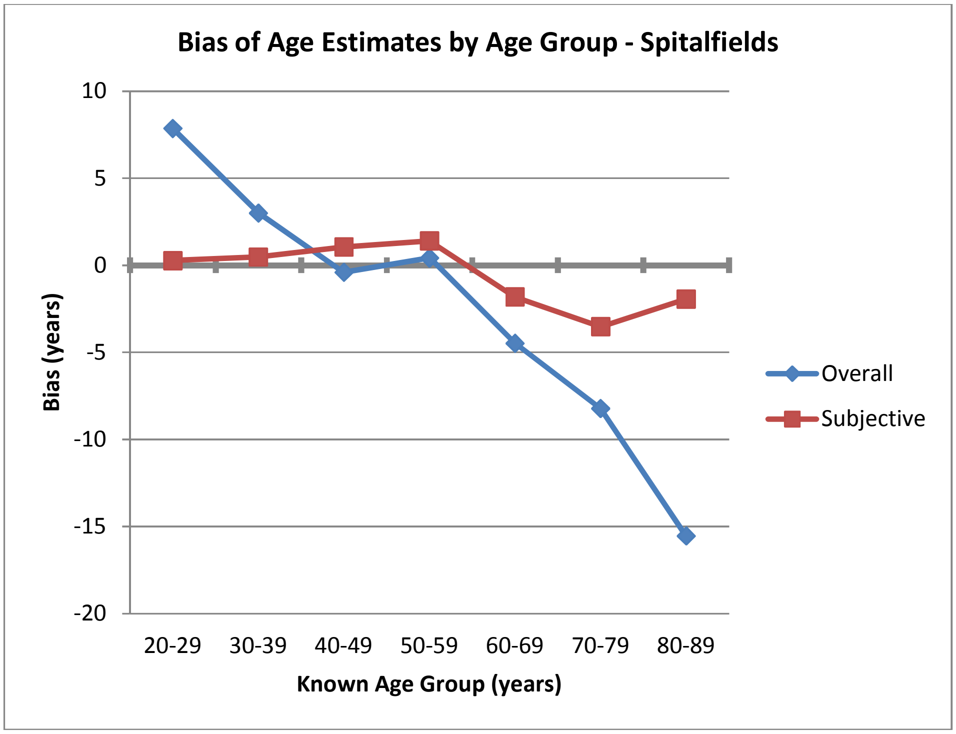


Figure 4.33. Bias of age estimates by age group for the Spitalfields Collection

4.14.3 Coimbra Collection

The percentage of correct age estimates using the overall method for the sexes pooled did not follow the parabolic pattern. Instead, they fluctuated over the youngest age groups slightly, increased for the 40 to 49 age group, and then dropped steadily until the 80 to 89 age group, where the correct estimates increased again (see Table 4.110).

The percentages of correct subjective age estimates did not follow the parabolic pattern (Table 4.111). Instead, the estimates generally decreased in percentage across the age range, beginning at 94.7% for the 20 to 29 year olds, and reaching as low as 25.0% for the 90 to 99 year olds. There were some minor fluctuations for the 30 to 39 and 40 to 49 year olds and the 70 to 79 and 80 to 89 year olds.

The inaccuracy, standard deviation and bias for the overall estimates for Coimbra followed the same pattern as described earlier, decreasing across the younger age groups before increasing steadily around the middle age groups (see Table 4.112 and Figures 4.34 and 4.35). Inaccuracy and standard deviation decreased until the 50 to 59 age group and increased from the 60 to 69 group onwards, with the exception of standard deviation for the 90 to 99 year olds. Bias again was positive for the younger age groups, but changed to increasingly negative values beginning with the 50 to 59 age group. For the subjective estimates, inaccuracy and standard deviation fluctuated, and bias generally increased from the youngest to oldest age groups, with some minor fluctuations, as with Spitalfields. Bias was positive for the three youngest age groups, and negative beginning from the 50 to 59 age group through to the oldest age groups. The highest amount of bias was -20.00 for the 90 to 99 age group using the overall method; the value for the subjective method for that age group was -4.50. The total mean bias, inaccuracy and standard deviation were lower for the subjective estimates than for the overall estimates.

Known Age Group	Number of Individuals	Correct Age Estimates (*n*)	Correct Age Estimates (%)
20-29	19	12	63.2
30-39	20	12	60.0
40-49	20	17	85.0
50-59	20	16	80.0
60-69	21	9	42.9
70-79	20	1	5.0
80-89	16	2	12.5
90-99	4	0	0.0
100+	--	--	--
Total	140	69	49.3

Table 4.110. Coimbra Collection, numbers and percentages of correct overall estimates

Known Age Group	Number of Individuals	Correct Age Estimates (*n*)	Correct Age Estimates (%)
20-29	19	18	94.7
30-39	20	15	75.0
40-49	20	16	80.0
50-59	20	13	65.0
60-69	21	14	66.7
70-79	20	9	45.0
80-89	16	8	50.0
90-99	4	1	25.0
100+	--	--	--
Total	140	94	67.1

Table 4.111. Coimbra Collection, numbers and percentages of correct subjective estimates

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	2.53	0.16	5.57	0.69	2.53	0.16
30-39	2.05	1.85	3.36	3.95	2.05	1.65
40-49	0.50	0.55	1.47	1.47	0.20	0.35
50-59	0.55	2.50	1.32	5.28	-0.05	-0.90
60-69	2.57	1.95	3.14	3.64	-2.57	-0.43
70-79	10.45	4.40	6.36	5.76	-10.45	-3.80
80-89	16.06	3.94	8.31	5.74	-16.06	-3.94
90-99	20.00	4.50	5.23	3.87	-20.00	-4.50
100+	--	--	--	--	--	--
Total (mean)	5.07	2.22	7.39	4.34	-3.63	-1.01

Table 4.112. Coimbra Collection, inaccuracy, standard deviation and bias for each age group, sexes pooled

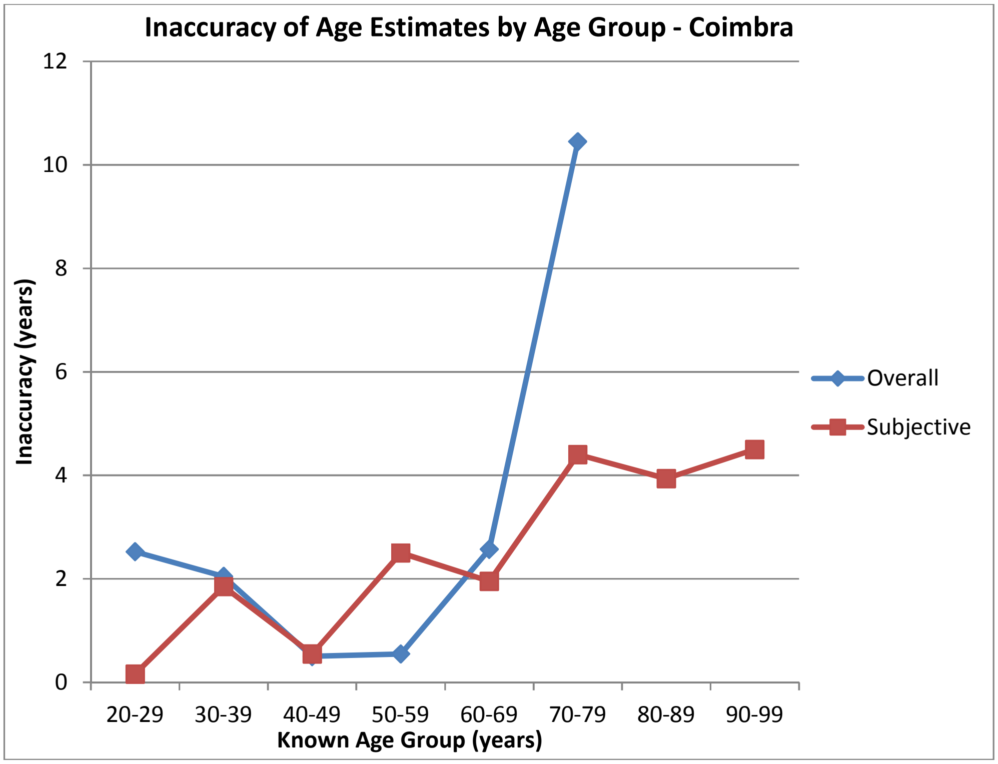


Figure 4.34. Inaccuracy of age estimates by age group for the Coimbra Collection

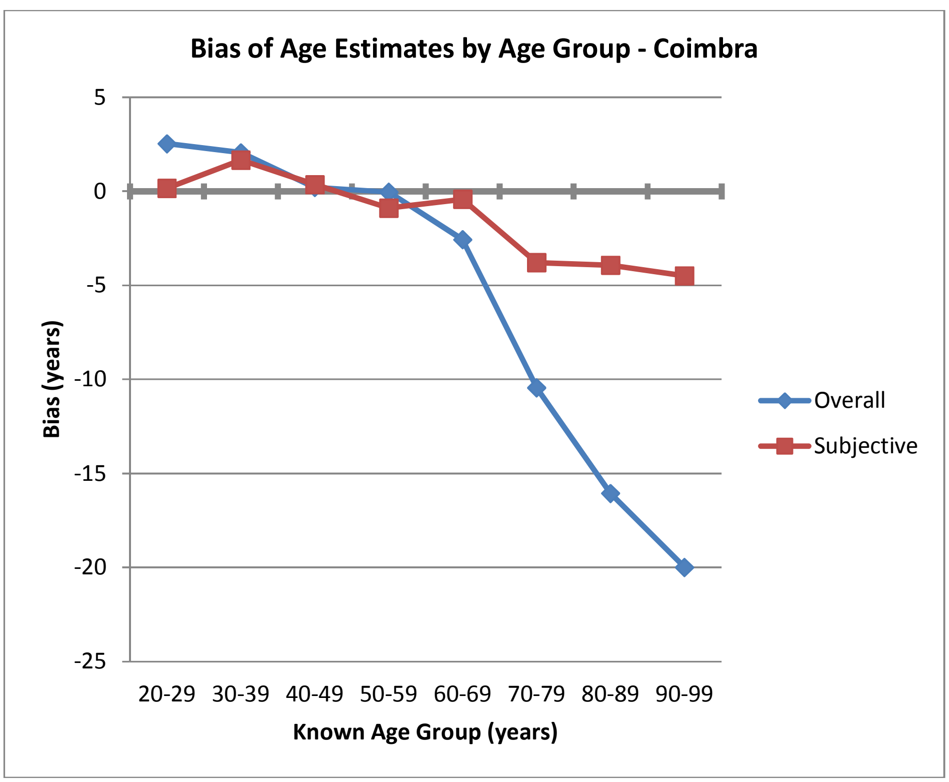


Figure 4.35. Bias of age estimates by age group for the Coimbra Collection

4.14.4 Lisbon Collection

The Lisbon Collection pattern of correct overall age estimates for the sexes pooled began by rising over the first three age groups, then consistently decreased from the 50 to 59 age group onwards (see Table 4.113). No individuals from the oldest age groups were correctly aged (from the 70 to 79 age group and older). The percentages of correct subjective age estimates did not follow the same pattern (see Table 4.114). Instead, the percentages decreased over the first three age groups before rising briefly at 50 to 59, then fluctuating somewhat over the remaining age groups, with lower proportions of correct estimates than those of the younger age groups. While none of the individuals aged 70 and over were aged correctly using overall estimates, 65.0% of the 70 to 79 year olds, and 55.0% of the 80 to 89 year olds were aged successfully using subjective estimates, as well as 14.3% of 90 to 99 year olds.

In terms of the overall method inaccuracy, the values steadily increased over the age range. Subjective inaccuracy fluctuated slightly over the age range, with a general increase from the youngest to the oldest age groups (Table 4.115, Figure 4.36). The overall method standard deviation fluctuated somewhat, but the general trend was one of increase over the age range; subjective method standard deviation also followed a general pattern of increase with age, with some fluctuations, including a surprising high of 10.76 for the 40 to 49 year olds. Excluding this peak, the range for subjective standard deviation was from 0.67 for the 20 to 29 age group to 7.87 for the 80 to 89 age group. The overall method bias also generally increased over the age range, with a slight fluctuation in the 30 to 39 and 40 to 49 age groups. Bias was positive for the first three age groups, changing to negative for the 50 to 59 year olds, and increasing consistently thereafter. Subjective bias was also positive for the first three age groups, and generally increased over the age range (Figure 4.37). The bias became negative for the 50 to 59 age group, as with the overall bias. Inaccuracy, standard deviation and bias were all lower for the subjective estimates than for the overall estimates.

Known Age Group	Number of Individuals	Correct Age Estimates (*n*)	Correct Age Estimates (%)
20-29	20	16	80.0
30-39	19	16	84.2
40-49	20	17	85.0
50-59	20	16	80.0
60-69	20	5	25.0
70-79	20	0	0.0
80-89	20	0	0.0
90-99	--	--	--
100+	--	--	--
Total	146	70	48.0

Table 4.113. Lisbon Collection, numbers and percentages of correct overall estimates

Known Age Group	Number of Individuals	Correct Age Estimates (*n*)	Correct Age Estimates (%)
20-29	20	19	95.0
30-39	19	18	94.7
40-49	20	14	70.0
50-59	20	17	85.0
60-69	20	10	50.0
70-79	20	13	65.0
80-89	20	11	55.0
90-99	--	--	--
100+	--	--	--
Total	146	103	70.6

Table 4.114. Lisbon Collection, numbers and percentages of correct subjective estimates

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	0.60	0.15	1.35	0.67	0.60	0.15
30-39	0.84	0.11	2.17	0.46	0.84	0.11
40-49	1.15	4.35	3.33	10.76	0.75	2.65
50-59	1.25	0.90	2.99	2.40	-1.25	-0.60
60-69	5.55	2.40	5.09	3.97	-5.55	-1.40
70-79	13.00	2.55	4.83	4.85	-13.00	-2.25
80-89	19.85	4.40	6.53	7.87	-19.85	-4.40
90-99	--	--	--	--	--	--
100+	--	--	--	--	--	--
Total (mean)	6.88	2.17	8.78	5.68	-6.24	-0.92

Table 4.115. Lisbon Collection, inaccuracy, standard deviation and bias for each age group, sexes pooled

4.14.5 Dart Collection

The percentages of correct overall age estimates for the Dart Collection increased over the first three age groups, before decreasing from the 50 to 59 age group onwards. The total percentage of correct overall age estimates was low, at 31.4% (see Table 4.116). The correct subjective estimates were fairly similar in proportions across the age range, only beginning to drop steadily from the 70 to 79 age group (Table 4.117). The total percentages of correct age estimations were higher for the subjective age estimates than for the overall age estimates.

Overall inaccuracy for the sexes pooled followed the pattern of initial decrease from the 20 to 29 group to the 40 to 49 age group, followed by a steady increase from the 50 to 59 group onwards. Overall standard deviation and bias followed the same pattern, although standard deviation had some minor fluctuations. Bias again began positively for the first three age groups, changing to negative for the 50 to 59 group, becoming higher in value until the oldest age group.

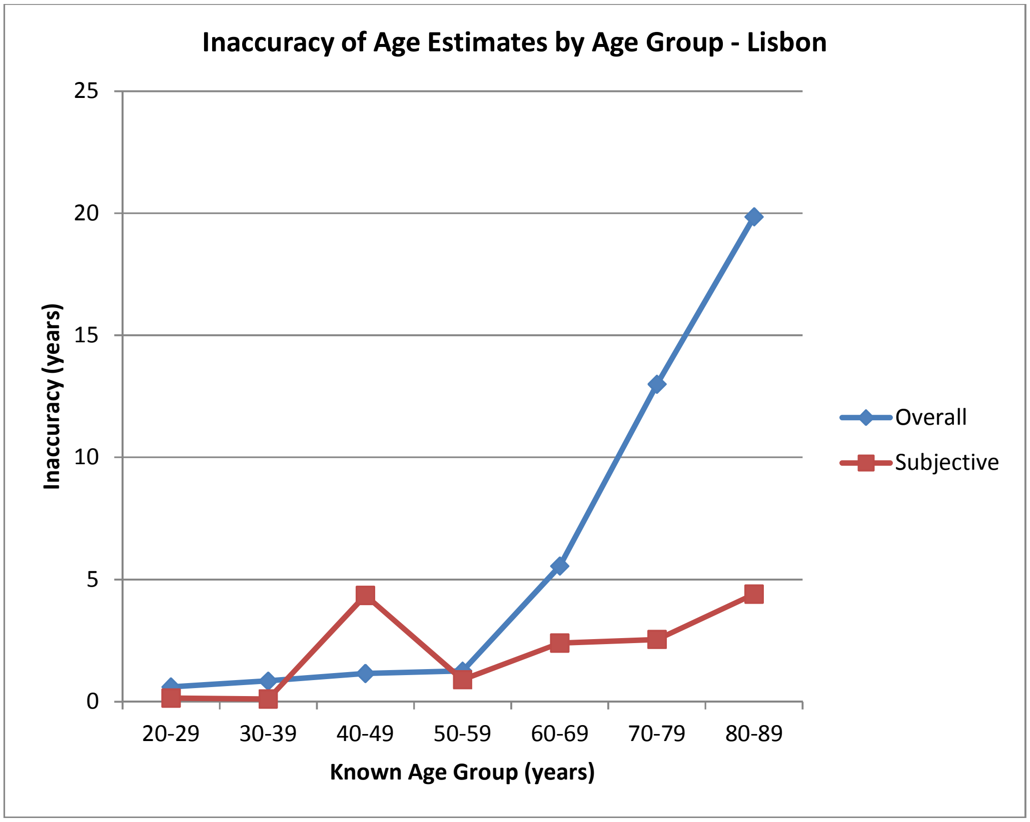


Figure 4.36. Inaccuracy of age estimates by age group for the Lisbon Collection

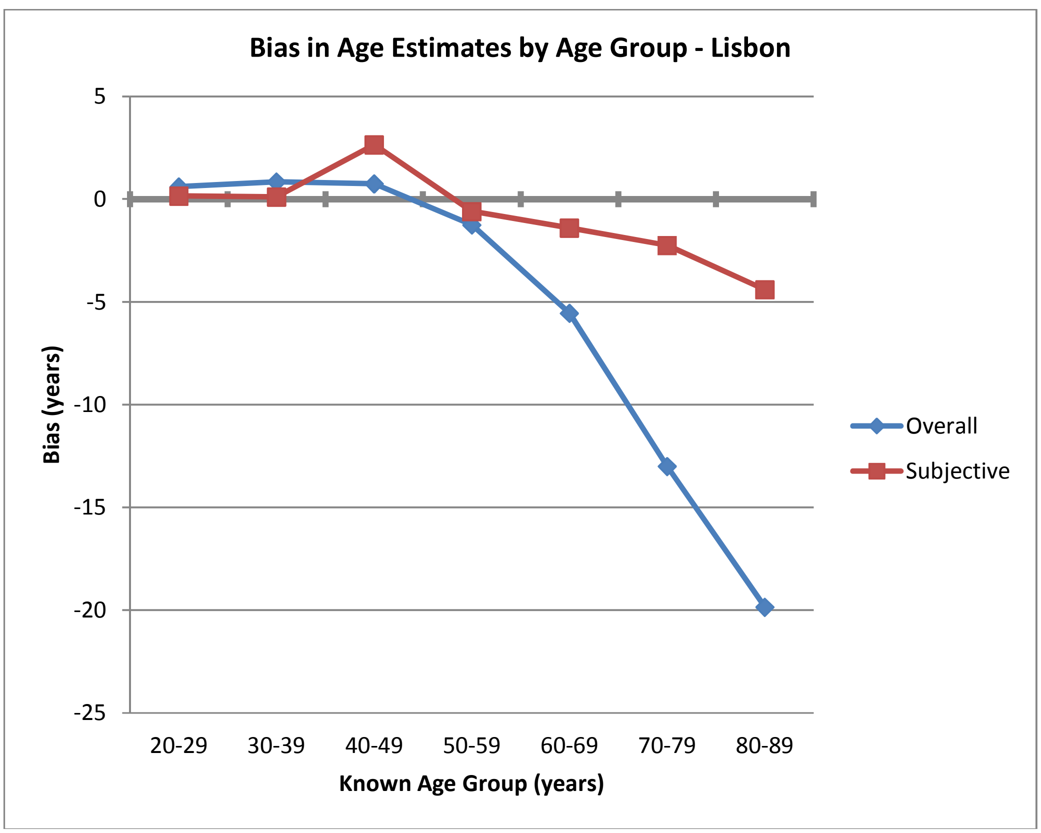


Figure 4.37. Bias of age estimates by age group for the Lisbon Collection

Subjective inaccuracy for the sexes pooled generally rose with age, apart from two minor fluctuations (at the 40 to 49 and 80 to 89 age groups). Subjective standard deviation fluctuated for the younger and middle age categories, but was lower for the 20 to 29 and 30 to 39 age groups, and highest for the two oldest age groups. Subjective bias fluctuated over the first three age groups, and then generally increased from 50 to 59 to the oldest age group. The first three age groups had a positive bias, while the age groups from 50 to 59 onwards had negative biases. Subjective inaccuracy, standard deviation and bias were lower than those of the overall estimates (see Table 4.118 and Figures 4.38 and 4.39).

Known Age Group	Number of Individuals	Correct Age Estimates (*n*)	Correct Age Estimates (%)
20-29	20	7	35.0
30-39	20	11	55.0
40-49	20	14	70.0
50-59	20	12	60.0
60-69	20	5	25.0
70-79	20	1	5.0
80-89	20	0	0.0
90-99	15	0	0.0
100+	4	0	0.0
Total	159	50	31.4

Table 4.116. Dart Collection, numbers and percentages of correct overall estimates

Known Age Group	Number of Individuals	Correct Age Estimates (*n*)	Correct Age Estimates (%)
20-29	20	13	65.0
30-39	20	11	55.0
40-49	20	13	65.0
50-59	20	13	65.0
60-69	20	13	65.0
70-79	20	11	55.0
80-89	20	9	45.0
90-99	15	4	26.7
100+	4	0	0.0
Total	159	87	54.7

Table 4.117. Dart Collection, numbers and percentages of correct subjective estimates

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	4.05	1.75	4.37	2.90	4.05	1.75
30-39	1.85	1.85	2.89	2.43	1.45	0.15
40-49	1.70	4.20	3.40	8.69	0.90	2.60
50-59	2.00	2.30	3.20	5.42	-1.80	-0.90
60-69	7.30	2.90	6.51	6.37	-7.30	-2.20
70-79	12.90	3.65	6.69	5.17	-12.90	-3.65
80-89	18.80	3.25	6.04	3.77	-18.80	-3.25
90-99	26.07	7.60	10.30	9.93	-26.07	-7.60
100+	48.00	28.00	4.40	10.42	-48.00	-28.00
Total (mean)	9.78	3.92	11.54	7.25	-7.99	-2.11

Table 4.118. Dart Collection, inaccuracy, standard deviation and bias for each age group, sexes pooled

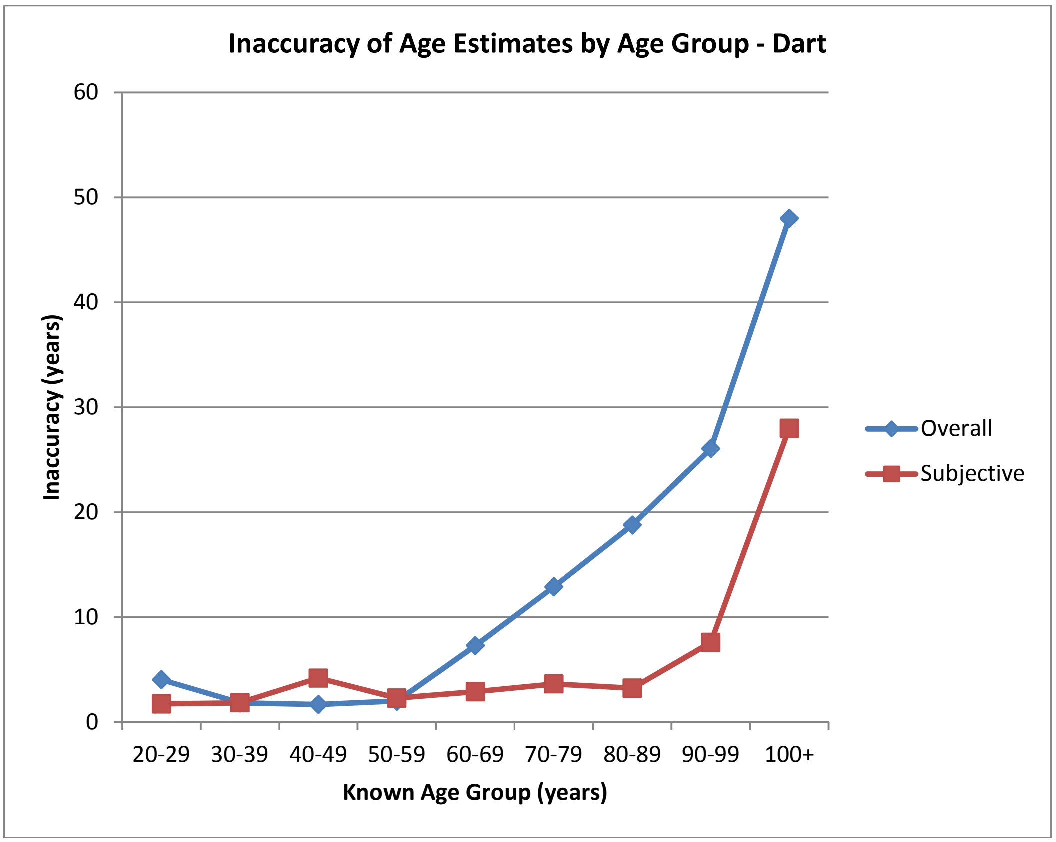


Figure 4.38. Inaccuracy of age estimates by age group for the Dart Collection

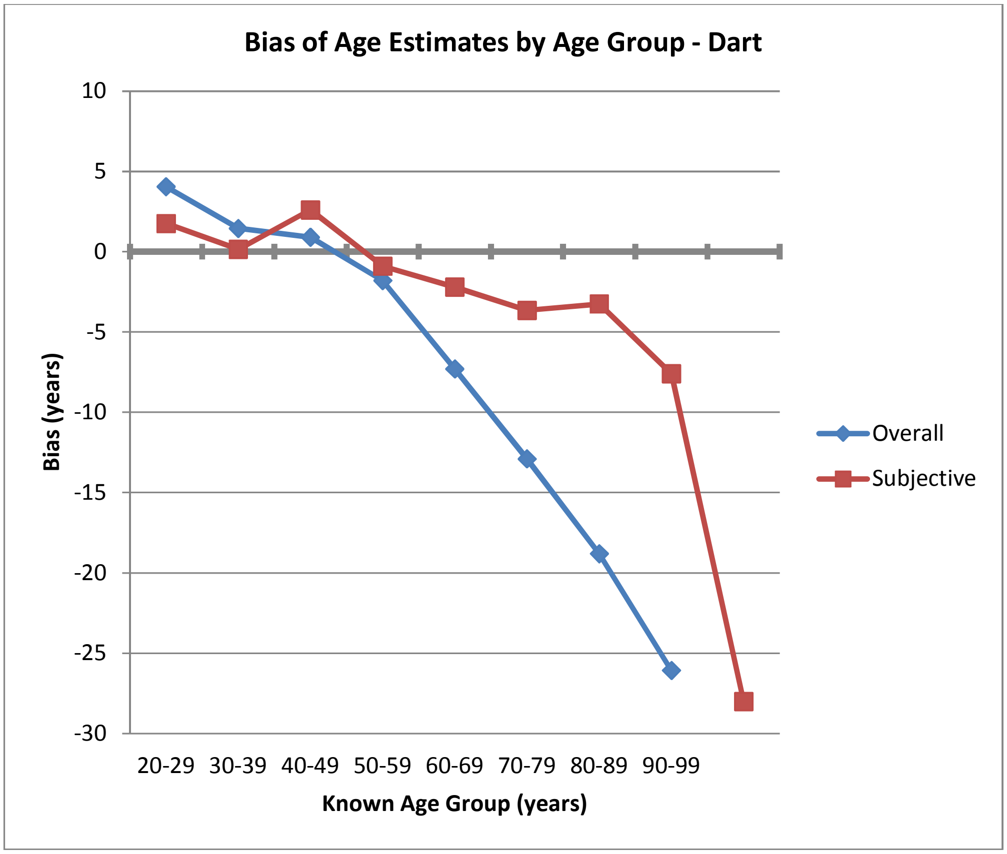


Figure 4.39. Bias of age estimates by age group for the Dart Collection

4.14.6 Pretoria Collection

For the Pretoria Collection, the percentage of correct overall estimates was fairly similar from the 20 to 29 through to the 50 to 59 age groups, decreasing thereafter. No individuals were correctly aged from 70 to 79 through to 90 to 99 years old (see Table 4.119). The percentages of correct subjective estimates fluctuated until the oldest age groups, then began decreasing from the 70 to 79 age group onwards (see Table 4.120).

The patterns for the overall inaccuracy, standard deviation and bias were the same as described elsewhere (parabolic), decreasing initially from the 20 to 29 age group until the 40 to 49 age group for overall inaccuracy and bias, and the 50 to 59 age group for overall standard deviation, increasing steadily thereafter, and peaking at the oldest age group (see Table 4.121 and Figures 4.40 and 4.41). Overall standard deviation was an exception to the peak at the oldest age group, as there was a drop at the 90 to 99 group compared to the 80 to 89 group. However, the 90 to 99 age group had only six individuals, compared to 20 individuals in the 80 to 89 age group, so perhaps this value was affected by the smaller group size.

Known Age Group	Number of Individuals	Correct Age Estimates (*n*)	Correct Age Estimates (%)
20-29	19	11	57.9
30-39	21	16	76.2
40-49	20	13	65.0
50-59	21	13	61.9
60-69	21	6	28.6
70-79	20	0	0.0
80-89	20	0	0.0
90-99	6	0	0.0
100+	--	--	--
Total	148	59	39.9

Table 4.119. Pretoria Collection, numbers and percentages of correct overall estimates

Known Age Group	Number of Individuals	Correct Age Estimates (*n*)	Correct Age Estimates (%)
20-29	19	15	79.0
30-39	21	17	81.0
40-49	20	16	80.0
50-59	21	15	71.4
60-69	21	11	52.4
70-79	20	15	75.0
80-89	20	9	45.0
90-99	6	3	50.0
100+	--	--	--
Total	148	101	68.2

Table 4.120. Pretoria Collection, numbers and percentages of correct subjective estimates

Subjective inaccuracy for the sexes pooled fluctuated over the age range, following no strict pattern; however, the oldest age groups did have the highest inaccuracy across the age range (3.75 and 4.50 for the 80 to 89 and 90 to 99 year olds, respectively). Subjective standard deviation also fluctuated across the age range. Subjective bias similarly fluctuated across the age range, although, again, the oldest age groups had the highest values. The overall bias was positive for the first three age groups, while subjective bias was positive for the first four age groups; the age groups following each of these all had negative biases. The highest amount of bias was -25.83 for the overall estimates for the 90 to 99 age group; the subjective bias for the same age group was -4.50. Total mean subjective inaccuracy, standard deviation and bias were lower than the same for the overall estimates.

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	3.32	1.05	5.15	2.72	3.32	1.05
30-39	1.62	2.33	3.57	7.56	1.14	2.14
40-49	1.35	1.10	3.05	2.36	1.05	1.10
50-59	1.48	0.81	2.86	1.63	-1.48	0.05
60-69	5.71	2.57	4.67	3.33	-5.71	-1.14
70-79	12.75	1.10	5.79	2.92	-12.75	-0.90
80-89	19.30	3.75	8.63	5.66	-19.30	-3.75
90-99	25.83	4.50	4.31	7.06	-25.83	-4.50
100+	--	--	--	--	--	--
Total (mean)	7.24	1.93	8.90	4.41	-5.67	-0.38

Table 4.121. Pretoria Collection, inaccuracy, standard deviation and bias for each age group, sexes pooled

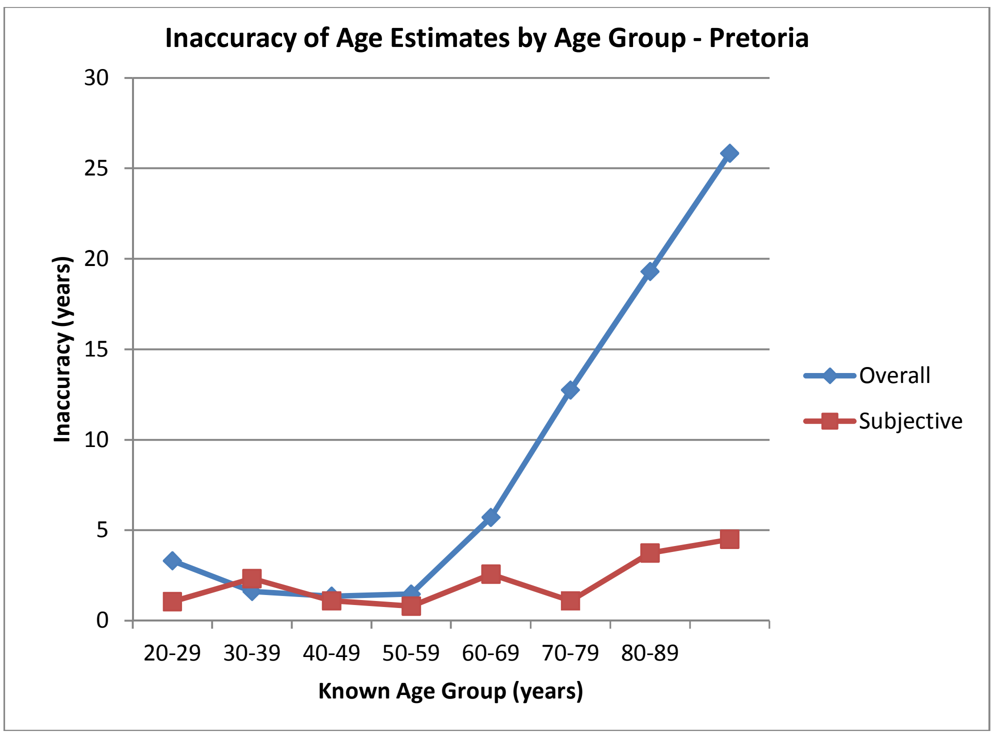


Figure 4.40. Inaccuracy of age estimates by age group for the Pretoria Collection

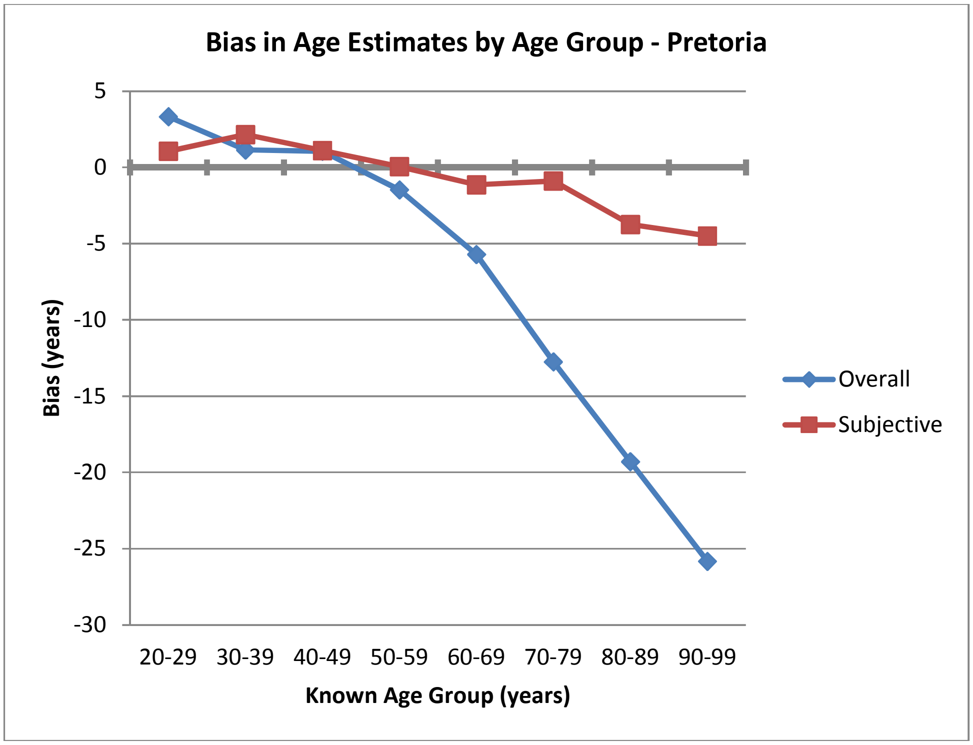


Figure 4.41. Bias of age estimates by age group for the Pretoria Collection

4.15 Subjective Skeletal Traits

As mentioned earlier, the subjective method of age estimation involved the use of the formal age estimation methods, alongside informal skeletal features that have not yet been studied extensively nor developed into age indicators with specific associated age ranges for use in bioarchaeology and forensic anthropology. Some of these have long been known, as noted in Chapter 2; others were discussed during a course at the University of Odense, as noted in Chapter 3. They are: new bone formation ("lipping") of the fovea capitis of the femoral head, spicules in the intertrochanteric fossa of the femur, and peaking at the sagittal suture with associated slight concavity (thinning and "scooping") of the parietal bones, "shingle-like" ribs, thinning of the maxilla, angularity of the lateral scapular borders, and joint degeneration and osteoarthritis. In this study, these were noted when present. Over the course of this study, it was also noticed that the fovea capitis "lipping" had some variants, including an "overspill" of the new bone formation onto the femoral head, porosity and/or "in-filling" of the fovea capitis; these were noted where possible to assess whether the variants might be age-related or simply be variations in the lipping. Similarly, the intertrochanteric fossae of femora of some individuals had spicules of bone of varying lengths, some not more than small nodules. The recording of these traits was not part of the original research plan, and so notes of absence, particularly for the Grant and Spitalfields collections (the first two locations for data collection), are not complete. This means that whether a lack of any of these traits was due to an actual absence of the trait, or absence or damage to the skeletal elements so that presence/absence could not always be determined. The general trends for the presence of these traits, plus that of osteophytes on joint margins (whose presence is age-related, at least in part; see, for example, Watanabe and Terazawa, 2006:159; Listi and Manhein, 2012: 1539 and Chapter 5) and osteoarthritis (again, whose presence is age-related, at least in part, although trauma can also be a predisposing factor; see, for example, Kirkwood, 1997: 683 and Chapter 5) are summarised here to analyse their utility in providing additional information for estimating age-at-death.

4.15.1 Fovea Capitis and Intertrochanteric Spicules/Osteophytes

The presence of fovea capitis lipping, porosity, overspill and/or in-fill was observed in every collection, although not in every skeleton; Table 4.122, below, summarises the numbers of individuals with fovea capitis lipping and intertrochanteric spicules.

4.15.1.1 Coimbra

For the Coimbra Collection, all females over 30 had lipping of the fovea capitis, with two exceptions, one 78 year old and one 84 year old. The reason for the lack of lipping was not clear for one of these individuals (either the femoral head was absent or the lipping was absent). Two 26 year old females also had some fovea capitis lipping. When present, overspill and in-filling were found in females aged 64 and over. The males were found to be slightly more variable in their expression of fovea capitis lipping: after age 41, most males had lipping (or porosity, etc.). Of the males over 41 years with no lipping, two were 42 years old, one was 57, one was 65, one was 75, and one was 78 years old. Five younger males also displayed fovea capitis lipping: one 29 year old, and four in their 30s. When present in males, overspill and/or in-fill occurred from age 41 onwards, although only one male of 41 displayed such expression; all others with overspill or in-fill were 54 and older. Intertrochanteric spicules or nodules were noted in only three Coimbra individuals, one aged 52, and two in their 60s.

4.15.1.2 Dart

In total, 130 Dart collection individuals were recorded as having fovea capitis lipping, porosity, in-filling or overspill. For Dart females over age 41 years, lipping occurred in all except three individuals: a 45 year old, a 70 year old and a 71 year old. However, four out of ten 25 to 29 year old females also had some lipping, as did four out of ten 30 to 39 year old females. When present, overspill and in-filling was more common after age 56, although one 42 year old female displayed some overspill. Dart males displayed fovea capitis changes from age 40 and over, with the exception of one 46 year old. Six young males also had fovea capitis lipping (two in their late 20s, and four in their 30s). When present in males, overspill and in-filling seems to occur at any age, but is very slightly more common at older ages.

Intertrochanteric spicules or nodules occurred in females from 44 years of age and older, although they were more common in females aged 55 and over. Three young males displayed intertrochanteric new bone formation (aged 31, 38 and 43), although it was more common in males aged 46 and over.

4.15.1.3 Grant

While there were 17 recorded instances of fovea capitis lipping, overspill, etc., for the Grant Collection, for most individuals, the presence or absence of the femora was not noted. This was the first collection from which data were collected, without systematic checking for the presence of fovea capitis lipping, so where no mention was made of lipping in the data notes, it is possible that lipping was absent or the element was missing. However, where fovea capitis lipping was present, it was mostly listed for individuals in their 60s and 70s; the youngest instance was in a 37 year old male. Three examples of intertrochanteric nodules (new bone formation) were noted for the Grant Collection, for individuals in their 60s, 70s and 80s.

4.15.1.4 Lisbon

For the Lisbon collection, 120 individuals were recorded as having fovea capitis lipping, overspill, or in-fill. All females had fovea capitis lipping of some kind from the age of 43 years onwards; five younger females had slight lipping (ages 27 and 28, and five females aged 30 to 39). With regards to females between the ages of 40 to 49 years, only one 42 year old did not have any lipping. Overspill and in-fill were slightly more common at older ages for females. All males aged 49 years and over had fovea capitis lipping of some kind, as well as the majority of males in their 40s (eight of ten aged 40 to 49 do have lipping). Some younger males display lipping: six of ten males aged 30 to 39 display slight lipping, as did a 26 and 27 year old male. When present, overspill and in-fill occurred in males aged 57 and older. Intertrochanteric nodules were observed in 77 individuals; while it was more common in those aged 47 and older, it was also observed in four younger individuals (aged 27, 35, 38, and two 43 year olds).

4.15.1.5 Pretoria

Fovea capitis lipping was also common in Pretoria individuals; 131 individuals were recorded as having lipping, overspill and/or in-fill. All females aged 40 years and older had lipping of some kind, as well as five of nine females aged 20 to 29 and eight of eleven females aged 30 to 39 (some of the younger individuals had only slight lipping). When present, overspill and in-fill were observed in one 35 year old and one 54 year old female, but were more common after age 60. All males aged 29 years and older had lipping, but four of ten males aged 20 to 29 years also had slight lipping. Overspill and in-fill, when present, was more common after age 50, although two males in their 30s and one 45 year old were observed to have some overspill. Intertrochanteric nodules were recorded in 66 individuals. One 35 year old female was recorded as having this new bone formation, but otherwise it occurred in females aged 45 years and older. For males, intertrochanteric nodules were more common in ages 60 and over, but were also found in three males aged 30 to 39, and four aged 40 to 49.

4.15.1.6 Spitalfields

Only 45 Spitalfields individuals were recorded as having fovea capitis lipping of some kind, but many Spitalfields skeletons were damaged, and notes on missing elements were not always complete. However, any lipping in females was only noted at age 43 and older, and while in-fill and overspill were noted in one 47 and one 57 year old female, they were more common after age 57. For males, one 27 year old had slight lipping of the fovea capitis; otherwise, lipping was present only in males aged 37 and older. Porosity and overspill were observed in males aged 58 years and older. Intertrochanteric nodules occurred in females aged 47 years or older, and in males aged 52 and older (although one 37 year old male had intertrochanteric nodules).

	Fovea Capitis		Intertrochanteric Fossa Spicules
	n: lip/overspill/etc.	*n*: no lipping	*n*: presence
Coimbra	109	16	3
Dart	130	26	74
Grant	17	?	3
Lisbon	120	26	77
Pretoria	131	12	66
Spitalfields	45	4?	21

Table 4.122. Numbers of individuals in each collection with presence/absence of fovea capitis lipping and present of intertrochanteric fossa spicules

4.15.2 Parietal Scooping, Thinned Maxillae, "Shingle-Like" Ribs and Angular Lateral Scapular Borders

Only one 75 year old Coimbra male was noted as having thinned parietal bones. Similarly, one 89 year old Dart female was recorded as having "possibly" thinned parietal bones. Six Lisbon individuals (two male, four female) were observed as having thinned and/or "scooped" parietal bones; these individuals were in their 70s, 80s and there was one 91 year old. Eight Pretoria females and nine Pretoria males were recorded as having scooped parietal bones, although one female was as young as 49; the descriptions of scooping in the younger males were marked with question marks, so these were likely representative of normal variation (without the thinning associated with age-related scooping of parietal bones).

One 93 year old Dart female was recorded as having very angular lateral scapular borders, and eight Dart females were recorded as having thinned maxillae: one 51 year old, two in their 70s, two in their 80s, and two in their 90s. Three Lisbon individuals were observed as having thinned maxillae (two females, aged 76 and 83, and one male, aged 67). Three Lisbon females were recorded as having thin or shingle-like ribs (aged 74, 88 and 92). Thinned maxillae were more common in Pretoria individuals after age 65, and one 90 year old Pretoria female was described as having shingle-like ribs. Three Spitalfields males were recorded as having thinned and sunken maxillae; one Spitalfields female was noted as having shingle-like ribs. Three Spitalfields males (aged 63, 66, and 76) had thinned and sunken maxillae, while one 86 year old female had shingle-like ribs.

4.15.3 Osteoarthritis and Marginal Osteophytes

Osteophytes on the margins of joints and osteoarthritic joints were recorded. The most common joint affected with OA was variable by collection. Vertebral osteophytes and/or OA were very common overall. The most common joints with OA are listed in Table 4.123, with numbers of affected individuals by collection; other uncommonly-affected joints were not listed in the table, but include one Coimbra individual with eburnation of the pubic symphysis, and one Dart individual with OA of the ankle.

	Joint or Skeletal Element with Osteoarthritis (number of individuals)						
	Shoulder	Sternal Clavicle	Elbow	Wrist	Hip	Knee	Spine
Coimbra	7	0	5	2	1	1	7
Dart	16	3	15	6	11	18	24
Grant	3	1	--	--	--	--	1
Lisbon	7	3	6	1	4	20	33
Pretoria	14	2	17	6	7	24	44
Spitalfields	13	10	1	2	12	4	8

Table 4.123. Numbers of individuals with OA in joints observed (joint components observed listed below). Shoulder: lateral clavicle, humeral head, and/or acromioclavicular joint; elbow: distal humerus and/or proximal ulna; wrist: distal radius and/or ulna; hip: acetabulum and/or femoral head; knee: distal femur and/or proximal tibia; spine: apophyseal joints.

No Coimbra females had osteophytes or OA from the ages of 20 to 29 years; only one Coimbra male had osteophytes in this same age range (vertebral). Three Dart females and three Dart males aged 20 to 29 displayed osteophytes; one of these males also was recorded as having

OA, but with possible underlying trauma. One Grant male aged 20 to 29 was recorded as having osteophytes, but no Grant females. No Lisbon or Spitalfields males or females aged 20 to 29 had osteophytes or OA. One Pretoria female and two Pretoria males aged 20 to 29 were recorded as having osteophytes.

Spitalfields, Lisbon, Grant, Dart females and Coimbra males aged 30 to 39 also had fairly low numbers of recorded instances of osteophytes; Pretoria, Dart males, and Coimbra females have higher numbers of 30 to 39 year olds with osteophytes.

For Spitalfields, Lisbon, Grant and Coimbra (Coimbra males in particular) osteophytes became more common and were observed in more joints per individual from the 50 to 59 age group onwards. For Pretoria and Dart, this occurs slightly earlier, from the late 30s and early 40s. The number of joints with osteophytes and OA tended to increase with age; OA became more common in ages 50 to 59 and 60 to 69. From the late 40s to 50s, vertebral osteophytes (the most common location across collections) in particular were very common, although slightly less so for Coimbra compared to the other collections. Osteophytes in more than one joint per individual became more common from the 40s onward, although for Lisbon males, this seems to happen in the 50s, and for Pretoria, this happens earlier, in the 30s. By the late 60s to early 70s and into the 80s and older, osteophytes and OA in multiple joints became more common.

4.16 Interobserver Error

Interobserver error testing was possible for some Spitalfields and Coimbra individuals, as Dr. Rebecca Gowland kindly offered her data from these collections for comparison, where data from the same individuals had been collected. The phases for Meindl-Lovejoy's auricular surface and Suchey-Brooks pubic symphysis were compared, as were the trait scores for Buckberry-Chamberlain's auricular surface method.

For the Meindl-Lovejoy auricular surface method, the numbers of individuals placed in the same phase by both Dr. Gowland and the author was fairly low for both Coimbra and Spitalfields. For Coimbra, only one right auricular surface (of 11) and nine (of 55) left auricular surfaces were placed in the same phase, while for Spitalfields, only three of 17 right auricular surfaces and 19 of 83 left auricular surfaces were placed in the same phase. The mean difference in phase placement for Coimbra was 1.5 phases, while for Spitalfields, it was 1.1 and 1.0 for right and left auricular surfaces, respectively. Table 4.124, below, has full details.

	Coimbra		Spitalfields	
	R	L	R	L
n compared	11	55	17	83
n M-L same	1	9	3	19
n M-L different	10	46	14	64
x M-L difference	1.50	1.52	1.09	1.02

Table 4.124. Meindl-Lovejoy interobserver error differences
R = right bone; L = left bone; M-L = Meindl-Lovejoy phase

For the Suchey-Brooks pubic symphysis method, the numbers of individuals placed in the same phase by Dr. Gowland and the author were higher than for the Meindl-Lovejoy method. For Coimbra, four of eight right and 23 of 50 left pubic symphyses were placed in the same phase, while for Spitalfields, five of 16 right and 22 of 56 left pubic symphyses were placed in the same phase; these values, except for right Spitalfields pubic symphyses, were at or near 50% agreement. The phase differences were, in all cases, less than one full phase on average; for Coimbra, the mean phase difference was 0.56 for right and 0.65 for left pubic symphyses and, for Spitalfields, the mean phase difference was 0.94 for right and 0.88 for left pubic symphyses (Table 4.125).

	Coimbra		Spitalfields	
	R	L	R	L
n compared	8	50	16	56
n S-B same	4	23	5	22
n S-B different	4	27	11	34
x S-B difference	0.56	0.65	0.94	0.88

Table 4.125. Suchey-Brooks interobserver error differences
R = right bone; L = left bone; S-B = Suchey-Brooks phase

The Buckberry-Chamberlain scores for each auricular surface characteristic were analysed individually; see Table 4.126 for Coimbra, and Table 4.127 for Spitalfields. For Coimbra, scores for left auricular surface transverse organisation, microporosity, macroporosity, apical change, and right auricular surface microporosity, surface texture and apical change agreed more often than disagreed (that is, agreed in over 50% of cases for these scores). The mean score difference for Coimbra was 1.00 (for right auricular surface transverse organisation) or less; the lowest mean score difference was 0.13 for surface texture scores of right auricular surfaces. For Spitalfields, left auricular surface scores for surface texture, microporosity, macroporosity and apical change, and right auricular surface scores for microporosity and macroporosity agreed more often than they disagreed – again, there was agreement in over 50% of cases for these scores. The mean score difference for Spitalfields had a narrower range than that for Coimbra, from a high of 0.89 for surface texture scores for right auricular surfaces to a low of 0.22 for microporosity scores (also for right auricular surfaces).

	Coimbra									
	R TO	L TO	R ST	L ST	R MI	L MI	R MA	L MA	R AP	L AP
n compared	8	57	8	55	8	54	8	54	9	55
n same	2	29	7	18	6	36	2	36	6	30
n different	6	28	1	37	2	18	6	18	3	25
x score difference	1.00	0.61	0.13	0.81	0.25	0.39	0.75	0.35	0.33	0.45

Table 4.126. Buckberry-Chamberlain component score interobserver error differences for Coimbra
R = right bone; L = left bone; TO = transverse organisation; ST = surface texture; MI = microporosity; MA = macroporosity; AP = apical changes

	Spitalfields									
	R TO	L TO	R ST	L ST	R MI	L MI	R MA	L MA	R AP	L AP
n compared	15	83	19	85	18	81	18	82	21	82
n same	7	41	6	43	14	61	10	54	9	51
n different	8	42	13	42	4	20	8	28	12	31
x score difference	0.53	0.52	0.89	0.59	0.22	0.28	0.56	0.37	0.55	0.37

Table 4.127. Buckberry-Chamberlain component score interobserver error differences for Spitalfields
R = right bone; L = left bone; TO = transverse organisation; ST = surface texture; MI = microporosity; MA = macroporosity; AP = apical changes

4.17 Intraobserver Error

An intraobserver error test was also undertaken, where 20% of the original Grant sample were re-recorded (Grant 2). The Grant Collection (Grant 1) was the first collection from which data were collected; all other data were collected from the other collections before Grant was revisited. This way, any error present should have been the maximum amount of error, as much experience was amassed between Grant visits, and approximately a year’s time had passed. Intraobserver error was analysed for the morphological pelvic traits, skull traits, and sex estimations using pelvis, skull, pelvis and skull combined, and the metrical method, for the Suchey-Brooks, Meindl-Lovejoy, and Buckberry-Chamberlain phase differences, as well as for the Buckberry-Chamberlain scored traits, and the inaccuracy and bias for the overall and subjective age estimates. Phase differences for the cranial suture methods, and for the sternal rib were not calculated as there were so few individuals with these skeletal elements available for observation.

4.17.1 Sex Determination Intraobserver Error

The agreement for the ventral arc and subpubic concavity were better than for the ischiopubic ramus ridge and greater sciatic notch. For ventral arc and subpubic concavity, agreement was as high as 90% and 85%, respectively, while for the ischiopubic ramus ridge and greater sciatic notch, it was only 60% and 35%, respectively (see Figure 4.42). The mean score differences were higher for the ventral arc and subpubic concavity, although the very low numbers of individuals used to calculate these means are the cause; for the ischiopubic ramus ridge, the mean score difference was 1.57, while for the greater sciatic notch, the mean score difference was 0.62 (see Table 4.128).

	Ventral arc	Subpubic concavity	Ischiopubic ramus ridge	Greater sciatic notch
n compared	20	20	20	20
n same	18	17	12	7
n different	2	3	8	13
x score difference	2.00	1.67	1.57	0.62

Table 4.128. Intraobserver error in morphological pelvic traits for sex determination, between Grant 1 and Grant 2

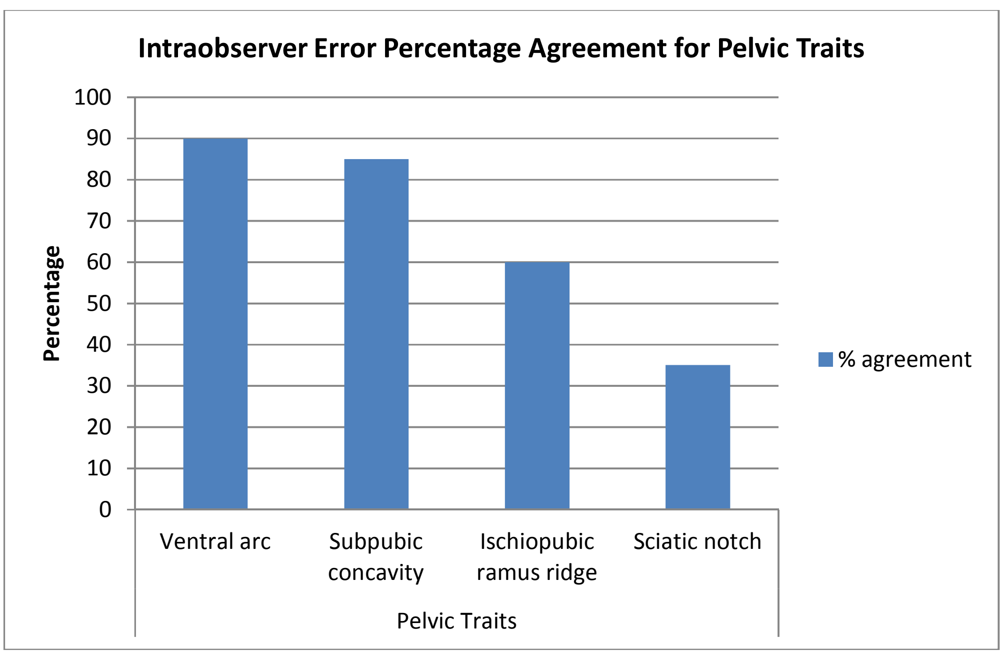


Figure 4.42. Bar chart of intraobserver percentage agreement for each pelvic trait

The agreement rate between the first and second Grant visits for the morphological traits of the skull used to determine sex were lower than for the pelvic traits. Agreement ranged from 44% for the mental eminence, to 41% for the glabella, followed by the nuchal crest, at 39%, 27% for the supraorbital margin, to a low of 25% for the mastoid process (see Figure 4.43). The mean score difference ranged from a high of 0.91 for the nuchal crest to 0.60 for both the glabella and mental eminence. Table 4.129 has full details.

	Nuchal crest	Mastoid process	Supraorbital margin	Glabella	Mental eminence
n compared	18	16	18	17	9
n same	7	4	5	7	4
n different	11	12	13	10	5
x score difference	0.91	0.75	0.81	0.60	0.60

Table 4.129. Intraobserver error in morphological skull traits for sex determination, between Grant 1 and Grant 2

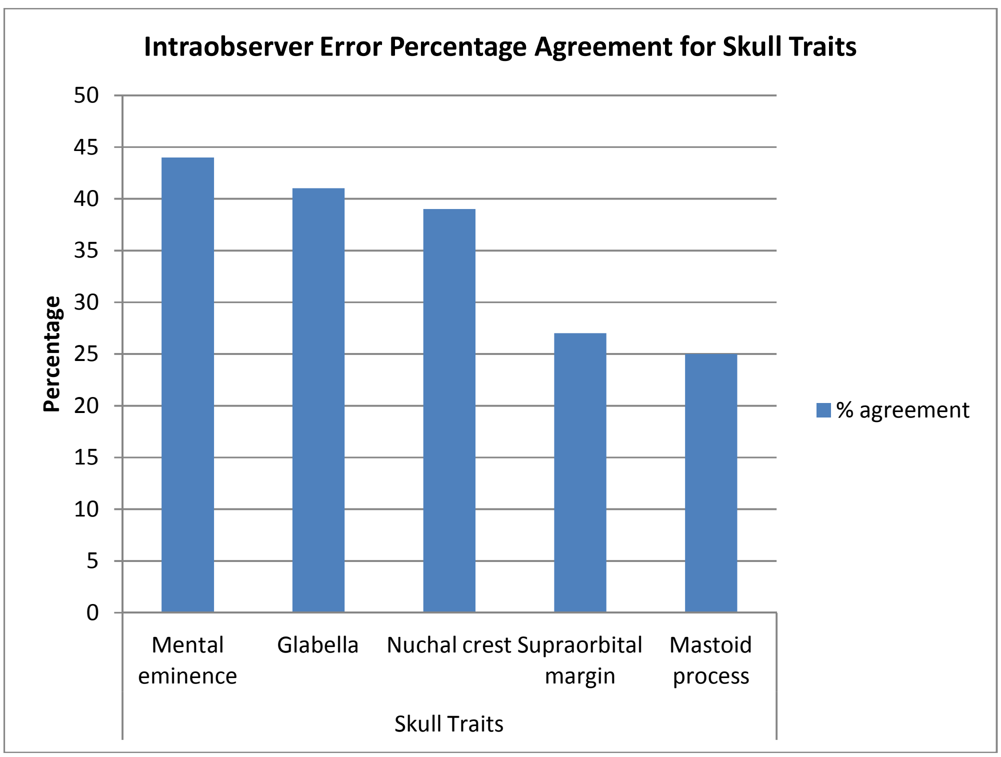


Figure 4.43. Bar chart of intraobserver percentage agreement for each skull trait

In terms of the sex determinations, agreement was generally good. For the sex determinations, only in one instance was the sex estimation actually changed from female to male (for the pelvic sex estimation); in all other cases, the difference was a matter of degree (for example, “probable female” to “female”). For sex determinations using the pelvis alone, there was agreement in 95% of cases (and a “score” difference of 2, for the one individual without agreement); for the skull alone, there was agreement in only 25% of cases, but with a “score” difference of 1; for the pelvis and skull combined, there was agreement in 90% of cases, with a mean “score” difference of 2. See Table 4.130, below, for absolute numbers.

	Pelvis	Skull	Pelvis+Skull
n compared	20	20	20
n same	19	5	18
n different	1	12	2
x score difference	2*	1	2

Table 4.130. Intraobserver error in sex determination, between Grant 1 and Grant 2
*only one individual, not a mean.

The metrical method was also analysed for intraobserver error. In terms of the estimated sex, there was agreement in all cases. The measurements themselves fared well, despite most having low agreement (0% to 28%): mean differences, measured in millimetres, ranged from a low of 0.23 mm for the maximum diameter of the femoral head, to a high of 2.91 mm for AIL. The percentage error was calculated for each measurement, and was generally low: 0.17% for the maximum length of the femur, 0.49% for the maximum diameter of the femoral head, 0.84% for the femur epicondylar breadth, 0.90% for the iliac breadth, 1.08% for SPRL, 1.12% for hip bone height, and 3.46% for AIL (see Table 4.131, below).

	Femur max length	Femoral head max diameter	Femur epicondylar breadth	Hip bone height	Iliac breadth	SPRL	AIL	Estimated sex
n compared	20	20	20	19	18	19	19	20
n same	5	0	3	4	5	0	0	20
n different	15	20	16	15	13	19	19	0
x difference (mm)	0.75	0.23	0.68	2.31	1.39	0.80	2.91	--
x difference with sign (mm)	0.70	0.09	-0.63	-1.92	-1.20	-0.56	-2.01	--
% error	0.17	0.49	0.84	1.12	0.90	1.08	3.46	--

Table 4.131. Intraobserver error in metric traits for sex determination, between Grant 1 and Grant 2

4.17.2 Age Estimation Intraobserver Error

The intraobserver error for the pubic symphysis (Suchey-Brooks) and auricular surface (Meindl-Lovejoy and Buckberry-Chamberlain) phases were analysed first. The Suchey-Brooks percentage of agreement was 54%, with a mean phase difference of 0.83; for Meindl-Lovejoy, the percentage of agreement was 44%, with a mean phase difference of 0.88; for Buckberry-Chamberlain, the percentage of agreement was 47%, with a mean phase difference of 1.13. Table 4.132, below, has details.

	Suchey-Brooks	Meindl-Lovejoy	Buckberry-Chamberlain
n compared	17	18	15
n same	6	8	7
n different	11	10	8
x phase difference	0.83	0.88	1.13

Table 4.132. Intraobserver error in age estimation, between Grant 1 and Grant 2

The Buckberry-Chamberlain scores were also analysed for intraobserver error. The highest agreement was for transverse organisation and apical change, both at 73%, followed by macroporosity, with an agreement of 60%. Surface texture and microporosity had the lower percentage of agreement, at 53% for each. The mean score difference was 1.00 for all features except microporosity, for which it was 1.14. Absolute numbers can be found in Table 4.133, below.

	Transverse organisation	Surface texture	Microporosity	Macroporosity	Apical change
n compared	15	15	15	15	15
n same	11	8	8	9	11
n different	4	7	7	6	4
x score difference	1.00	1.00	1.14	1.00	1.00

Table 4.133. Intraobserver error in Buckberry-Chamberlain score between Grant 1 and Grant 2

Finally, the differences in inaccuracy and bias for the overall and subjective age estimates were compared between the first and second Grant visits (Table 4.134). For the overall age estimates, accuracy improved in the second Grant visit, from an inaccuracy value of 6.82 for the first visit to 6.12 for the second. Bias for the overall estimates remained the same, at -4.24 each time. Both subjective inaccuracy and bias were improved on in the second Grant visit; inaccuracy for the first round of data collection was 6.59, compared to 1.71 for the intraobserver round of data collection, while bias for the first visit was -4.82, and bias for the second visit was -1.12.

Overall Age Estimates				Subjective Age Estimates			
Inaccuracy G1	Inaccuracy G2	Bias G1	Bias G2	Inaccuracy G1	Inaccuracy G2	Bias G1	Bias G2
6.82	6.12	-4.24	-4.24	6.59	1.71	-4.82	-1.12

Table 4.134. Intraobserver error in overall and subjective age estimates between Grant 1 and Grant 2.
G1: Grant 1; G2: Grant 2

4.18 Conclusion

This chapter has presented the results of the comparisons and statistical tests used to analyse the sex and age determination methods used in this thesis to look for variability in human rates of ageing and sexual dimorphism, and the results of error testing. Key findings from the results included significant differences in distribution and median of sexually dimorphic traits between collections (the glabella, for example) and in phase or score distribution and median of ageing methods between collections (the highest phases of the auricular surface methods, for example). Mean ages per phase were found to significantly differ between collections for some ageing methods. In some cases, collections from the same country were found to have more similar results compared to collections from other countries, although some significant differences were found within the South African and Portuguese collections. Furthermore, the subjective method of age estimation offered improved values of accuracy, standard deviation and bias compared to age estimates based on the overall method. The following chapter will discuss these results with regard to their relationship to ageing rates and scales of sexual dimorphism, and their significance for bioarchaeological and forensic anthropological research.

Chapter 5: Discussion

5.1 Introduction

This chapter discusses the bioarchaeological and forensic anthropological implications of the results described in Chapter 4. The results for each age and sex estimation method will be discussed in turn, particularly results suggesting variation in ageing rates and sexual dimorphism in the different samples. The wider implications will be outlined following each section. Sexual dimorphism and ageing rates are discussed in the context of a scale, or continuum, of variation; sexual dimorphism in terms of more “male” or more “female” ends of a spectrum, and ageing rates in terms of relatively higher or lower ageing rates, or ageing more quickly or more slowly.

5.2 Limitations

5.2.1 The Concept of Age

The first, and perhaps most important, issue relating to this research is conceptual: physiological age, which is what is being evaluated with skeletal indicators, is not the same as chronological age, which is the information that bioarchaeologists are trying to access. While there is no solution for this per se, the use of multiple indicators may help to mitigate this issue.

5.2.2 Reliability of Known Ages of Skeletons in Documented Collections

The second limitation was with the reliability of the “known” ages. As discussed in Chapter 3, other researchers have noted problems with the reported ages for the Dart Collection (Tal and Tau, 1983: 217; Dayal et al., 2009: 8). This has the potential to affect any known age collection and the research carried out on it, particularly for collections that rely on unclaimed bodies for their skeletal material, or in societies where exact ages are not always known. For the Dart Collection, for instance, sometimes, when individuals died in hospital with no family present, their ages at death were estimated by the doctors or hospital staff. As outlined in Chapter 3, it is possible to test for unreliability of ages by looking for age heaping – the tendency to report particular terminal digits in stated ages (and the corresponding avoidance of other terminal digits); 0 and 5 tend to be frequent. Bar charts were constructed to look for age heaping for all of the collections sampled in this study; however, full collection lists were not available for Coimbra or Grant, so only lists of the individuals sampled here were used to construct the bar charts. The full collection lists were more helpful in assessing reliability of ages, but as the Coimbra and Grant samples were randomly chosen, they are hopefully representative of the ages in the collections.

Age heaping was found in both the Dart and Pretoria Collections. The Dart results were not surprising, given the previously published results of Tal and Tau (1983) and Dayal et al. (2009). Age heaping was found for all ethnicities and both sexes, although it was most pronounced in non-white ethnicities; the pooled sexes bar chart is presented in Figure 5.1. Interestingly, the age heaping found in documented ages from the Pretoria Collection seem to be confined to individuals with ethnicities listed as black or 'other' ('other' refers to "coloured" individuals in the Pretoria Collection and individuals listed as 'hybrid' or 'mixed' in the Dart Collection). White individuals do not show evidence of age heaping. These differences are not constrained by sex; males and females listed as black or 'other' exhibit age heaping, while males and females listed as white do not. It was also possible to divide the Pretoria Collection sample into during- and post-apartheid years, to see whether the racist attitudes of apartheid-era South Africa affected the assessment of age of individuals dying in hospital by staff. However, no differences were noted when the collection was divided by those who died before 1994 and those who died from 1994 to the present. Of course, racism may not have been a characteristic of all staff who estimated ages-at-death for individuals dying in hospital.

Similarly, the end of apartheid did not necessarily signal the end of racist attitudes in those who held them. In any case, individuals listed as 'black' or 'other' have less reliable ages-at-death both pre- and post-1994, while individuals listed as 'white' seem to have reliable documented ages at death both pre- and post-1994. The bar graph for the sexes pooled for the Pretoria Collection is presented below; particularly high frequencies of 'black and other' individuals can be seen for ages 60 and 70 (Figure 5.2). None of the other collections show any indication of age heaping; of course, the numbers of individuals dying at any particular age vary, and there are some "peaks", but not in large numbers, or corresponding to any particular sets of terminal digits. To avoid potentially unreliable ages in the Dart and Pretoria Collections, individuals with stated ages-at-death ending in 0 or 5 were avoided as much as possible. Dayal et al. (2009: 8) also suggested this strategy for research involving documented collections with potentially unreliable documented ages. The bar charts for the other collections, where no evidence for age heaping was found, are in Appendix 10.

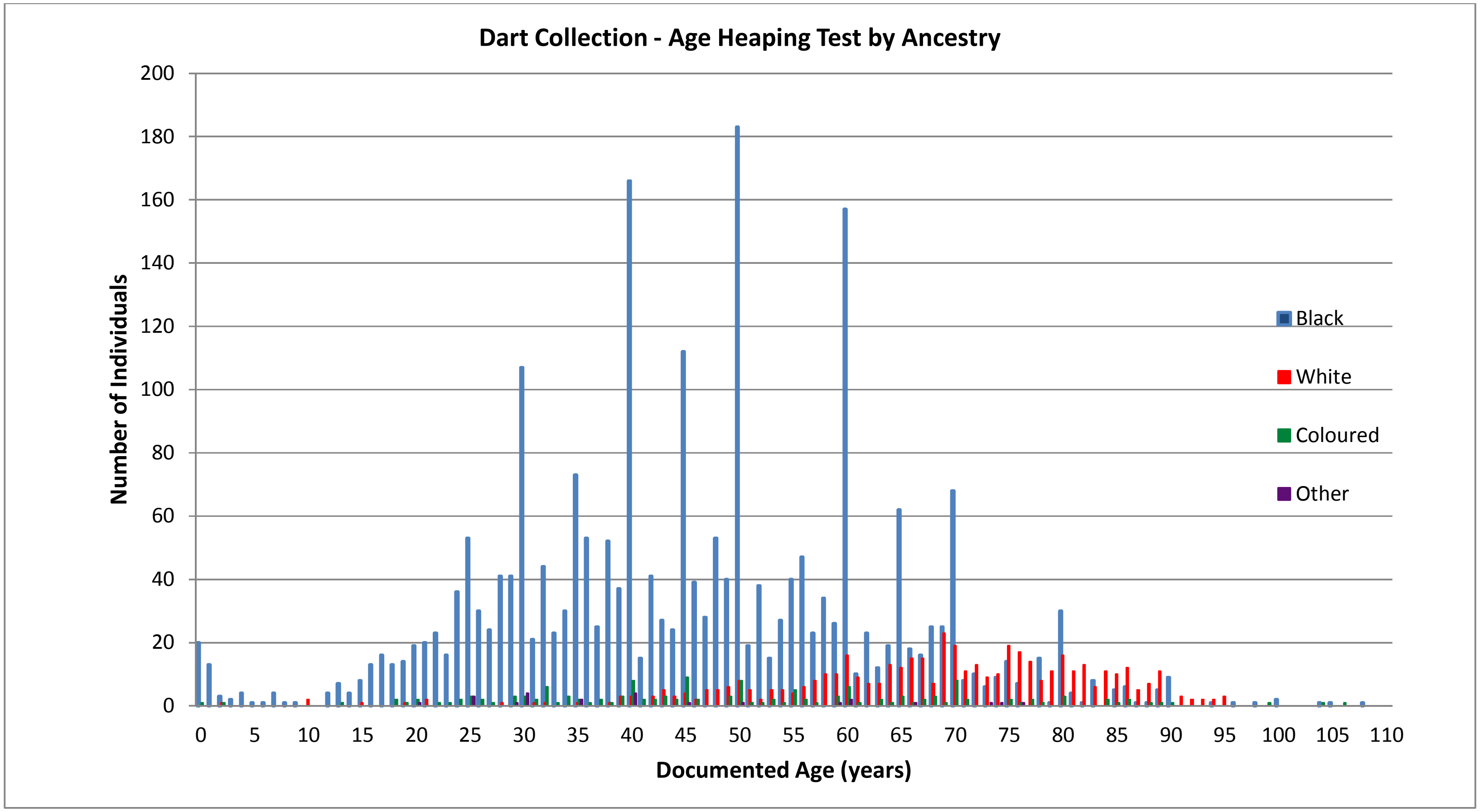


Figure 5.1. Documented Dart Collection ages-at-death, divided by ancestry, showing age heaping.

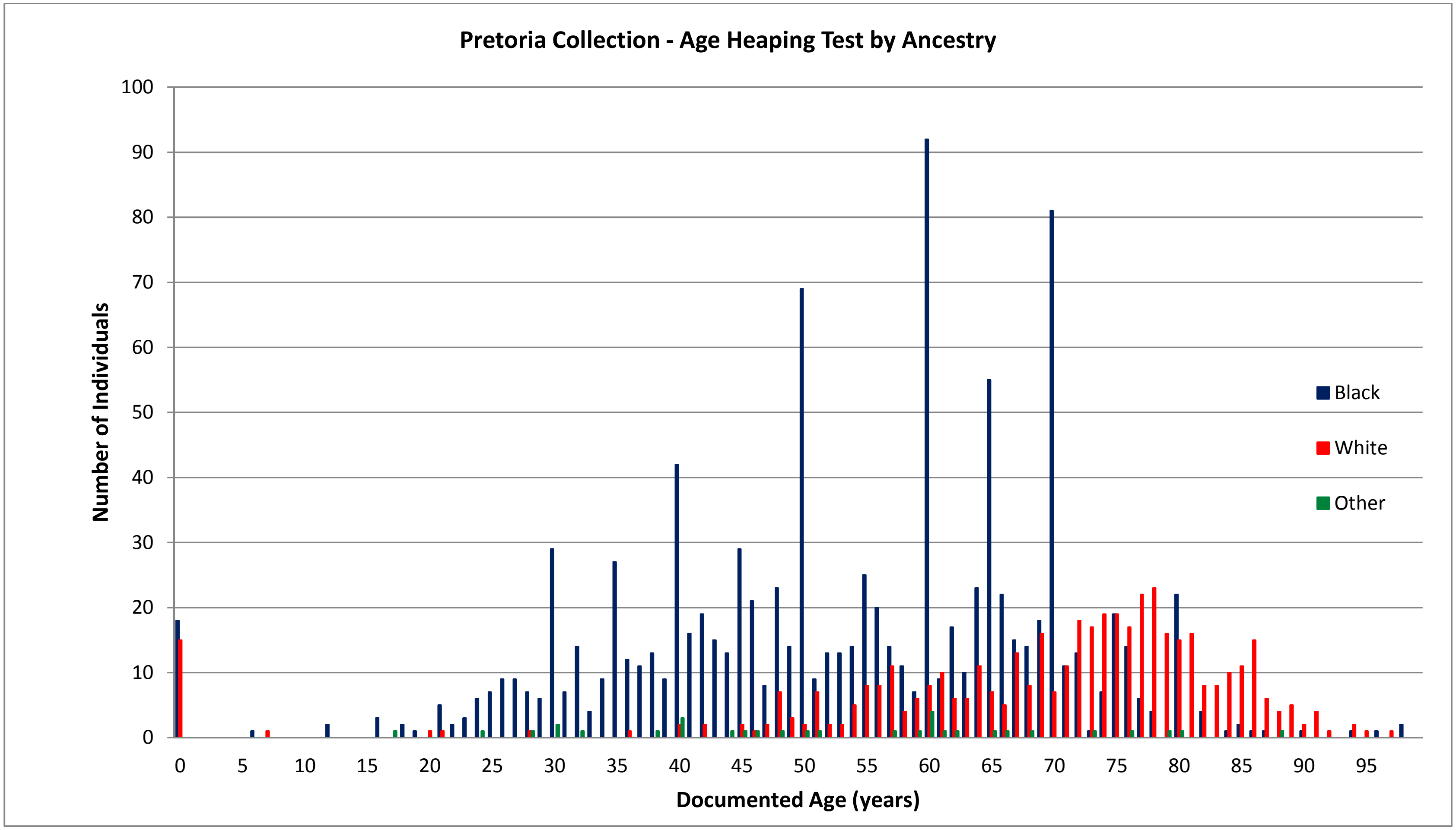


Figure 5.2. Documented Pretoria Collection ages-at-death, divided by ancestry, showing age heaping.

5.2.3 Numbers of People in Age Groups

The third limitation – related to the Grant Collection – is that the small number of females in the collection did not allow for the desired ten females for each age-at-death decade. All females were studied, but the number was still low, with a total of 18 females. As a result, the age distribution is not uniform. This is problematic for the pooled results, as in many instances (described in more detail below) significant differences that were found between the Grant Collection and the other collections were the result of the low female sample size of the former, not reflective of morphological differences. The only solution here was to be aware that the low sample size was potentially problematic, and note when the non-uniform age structure, rather than real differences between collections, was the reason for significant differences. The results were also analysed by sex, which also helped mitigate this problem, as proportions of Grant females compared to females of other collections should be comparable.

5.2.4 Labelling of the Coimbra Collection

Another possible limitation was noted in the Materials and Methods chapter, that the skeleton numbers on boxes of skeletons were colour-coded by sex for the Coimbra Collection. However, this was found not to have biased the results – that is, allocation accuracies for Coimbra sex determinations are not significantly higher than those of the other collections.

5.2.5 Limited Time Range for the Collections

The limited range of the time periods represented by the samples may also be a limitation of this study, because archaeological populations obviously range very widely across time periods. It would be useful to analyse the differences in ageing rate and expression of sexual dimorphism of an earlier population (medieval, for instance), in order to test whether such differences increase as increasingly older populations are considered. However, no documented and accessible collection earlier in date than Spitalfields is known.

5.3 Sex Determination: Pelvis and Skull Traits and Variability in Expression of Sexual Dimorphism

The Walker and Phenice methods using visual assessment of morphological sex differences require scoring of parts of the skull (glabella, supraorbital margin, mastoid process, nuchal crest, and mental eminence) and the pelvis (sciatic notch, ischiopubic ramus ridge, subpubic concavity and ventral arc). The statistically significant results of the analysis for each feature for the skull and pelvis will be discussed in turn.

The Kolomogorov-Smirnov tests indicated significant differences in the distribution of scores for all traits between the Grant Collection and the other collections when sexes were pooled. However, this is indicative of the paucity of female skeletons in the Grant Collection, rather than a real difference in the shape of the distribution (for example, a distribution may be flatter and wider, or more highly peaked) of sexual dimorphism, as the K-S test analyses the cumulative distribution. When the distributions of skeletal features were analysed by sex, the results reflected any differences in score distribution shape, rather than sampling issues.

5.3.1 Glabella (Figures 4.1a to 4.1f, Chapter 4)

No female score distributions were significantly different from that of any other collection, and no age-related trends were seen in either females or males. However, for males only, South African males showed a significant difference compared to males from all other collections, in both score distribution and central tendency. The Dart and Pretoria males have lower glabella scores, with more morphologically female and undetermined scores compared to the other collections. This suggests that the morphological expression of the glabella in South African males is skewed towards the “female” end of the spectrum. However, the female scores showed no significant differences between any collections. This suggests that the glabella is simply not as sexually dimorphic for the South African collections compared to the other collections, and the entire distribution is skewed towards the lower, morphologically “female” scores. This is supported by the relatively high allocation accuracy for females using the glabella (83% for both Pretoria and Dart) and low allocation accuracy for males (27% and 29% for Dart and Pretoria, respectively).

The glabella (and mastoid process) score distributions for Lisbon and Coimbra and for Dart and Pretoria males seem to resemble each other more closely than they do any of the other collections’ score distributions, suggesting some geographic clustering of likeness in sexual dimorphism. This provides evidence for close geographic proximity resulting in similarity in scale of sexual dimorphism.

5.3.2 Supraorbital Margin (Figures 4.2a to 4.2f, Chapter 4)

No age-related trends were observed for males or females. The Coimbra female scores for supraorbital margin are shifted slightly (statistically significant) towards the male end of the spectrum when compared to Lisbon and Dart females. It is interesting that significant differences were found in score distribution and central tendency between Coimbra and Lisbon, as these collections, and the origins of most of the individuals found in these collections, are geographically not far apart. Other similarities between the collections were discussed in Chapter 3, and include ethnocultural and linguistic similarities, and likely similar diets and socioeconomic status; both are urban localities (although Coimbra is a smaller city). That significant differences can be found within a country for some morphological skeletal features indicative of sexual dimorphism is important to note.

The scores of Grant males are centred more heavily on the “male” side of the spectrum compared to other collections; indeed, the Grant Collection as a whole is skewed towards the male, as the majority of Grant females scored 3 (“undetermined”). Meanwhile, the pooled score distributions for Lisbon and Dart are not obviously bimodal, suggesting lower levels of sexual dimorphism of the supraorbital margin compared to the other collections. However, the score distributions for all of the collections tend to focus on scores of 3 for both males and females; sexual dimorphism is not terribly high in general for the supraorbital margin.

5.3.3 Mastoid Process (Figures 4.3a to 4.3f, Chapter 4)

No age-related trends were seen in mastoid process score distributions.

Lisbon females’ higher proportions of lower scores reflect Lisbon females more highly “female” morphological expression of the mastoid process compared to the other collections. The Spitalfields female score values are more evenly spread, while the other collections’ score distributions are more highly peaked; the male Spitalfields scores are also fairly broad in range. The male score distributions for Lisbon and Coimbra are very similar to each other, as are the male score distributions for Dart and Pretoria. Dart and Pretoria males’ score values are skewed towards the “undetermined”/”female” side of the spectrum relative to Grant, Coimbra, Lisbon and, to some extent, Spitalfields. The within-country clustering of results here is interesting, and in direct contrast to the significant differences found between Lisbon and Coimbra for the supraorbital margin. This evidence makes it clear that while some traits may cluster in terms of expression of sexual dimorphism between populations that are close geographically, other traits may exhibit significantly different ranges of expression between the same two populations.

The Coimbra and Lisbon score distributions are more clearly bimodal than that of the other collections, and thus, more sexually dimorphic in mastoid process morphological expression. The score distributions for Dart and Pretoria are less clearly bimodal, with an overall shift to the “undetermined” and “female” end of the spectrum. This suggests that the mastoid process exhibits lower levels of sexual dimorphism in the South African collections compared to the other collections. Sexual dimorphism is higher for the mastoid process than for the supraorbital margin; the scores are more clearly bimodally distributed in most collections, rather than having a tendency to heap around the middle scores.

5.3.4 Nuchal Crest (Figures 4.4a to 4.4f, Chapter 4)

There were significant differences between Dart and Lisbon and Coimbra and Lisbon. The Dart score distribution is heavily skewed towards the female side of the spectrum, as both males and females tend to have "female" scores. Bimodality is not clear; thus, Dart exhibits lower levels of sexual dimorphism in the nuchal crest compared to the other collections. Lisbon has a bimodal distribution, but has some heaping of males and females, in particular, for the "undetermined" score. Coimbra, meanwhile, has a bimodal score distribution, but less clearly than does Lisbon. More Coimbra females have "female" scores of 1. Compared to Lisbon, then, the Coimbra score distribution is slightly skewed towards the "female" end of the spectrum.

For females only, Coimbra is different from all other collections, except Spitalfields. This is due to Coimbra females being more "female" in nuchal crest morphology compared to other collections, suggesting that Coimbra is slightly more highly sexually dimorphic for this trait. Pretoria and Spitalfields females also showed significant differences in score distribution, as Pretoria females are skewed towards the "male" side of the morphological spectrum compared to Spitalfields. In terms of males, Grant and Lisbon exhibit nuchal crest morphology skewed more heavily towards the higher "male" scores compared to Coimbra, Dart, and Pretoria. Again of note here are the significant differences between Lisbon and Coimbra; even over short geographical distances, significant variation in the scale of sexual dimorphism is possible.

5.3.5 Mental Eminence (Figures 4.5a to 4.5f, Chapter 4)

No age-related trends were seen in the results, and no significant differences in any score distributions were found. The female score distributions, in particular, have a very similar appearance. While the score distributions are bimodal, with a female peak at scores of 2, and a male peak around scores of 3, the mental eminence does not seem to be a highly sexually dimorphic trait. However, Dart and Pretoria males have scores that are slightly skewed towards the "female" end of the spectrum compared to the other collections.

5.3.6 Sciatic Notch (Figures 4.6a to 4.6f, Chapter 4)

No significant differences were found between females, although Dart's female score distribution was slightly wider and flatter than that of the other collections. This suggests that Dart females are more highly variable in sciatic notch expression compared to females of the other collections. The differences found in Lisbon males' score distribution compared to Dart, Grant and Spitalfields suggest that for the sciatic notch, Lisbon's male morphology is relatively more "female" than that of the other collections. The Spitalfields, Dart and Grant distributions are slightly wider and flatter than the distributions of Coimbra, Lisbon and Pretoria, suggesting that variation in male sciatic notch morphology is slightly more common for the former group of collections. Thus, the usefulness of the sciatic notch in determination of sex is somewhat dependent on the collection or population being analysed.

5.3.7 Ischiopubic Ramus Ridge (Figures 4.7a to 4.7f, Chapter 4)

The differences between Pretoria and the other collections for the sexes pooled are interesting. The differences are because Pretoria's score distribution is skewed slightly towards the female end of the spectrum of sexual dimorphism. For females only, the differences remain between Pretoria compared to Coimbra and Lisbon and, for males only, between Pretoria compared to Coimbra, Dart and Grant. The female differences are because Pretoria's female score distribution is skewed towards the morphologically "female" compared to the other collections. This provides more evidence for the possibility of significant differences within countries, with only short distances separating populations geographically. Pretoria males, meanwhile, are somewhat unusual (compared to the other collections) in that, despite a high frequency of "male" scores, there is another peak in proportions of males with "probable female" scores. The Grant score distribution is skewed more highly towards the morphologically "male" end of the spectrum compared to the other collections (significantly different from Lisbon only). The ischiopubic ramus ridge is more clearly sexually dimorphic, with a bimodal distribution, compared to the single skull features.

It also seems interesting that there are no Grant males with scores of 3, no Coimbra males with scores of 2, and no Dart males with scores of 2 or 4, given that in each of these cases, there are males with the surrounding scores. This is perhaps an artifact of the original scoring system using words instead of numbers – "female", "probable female", "undetermined", "probable male" and "male", which was converted to a numeric format after data collection for ease of analysis. Perhaps the inclination is to use the more definite "female" and "male" scores whenever possible.

5.3.8 Subpubic Concavity (Figures 4.8a to 4.8f, Chapter 4)

The subpubic concavity is highly sexually dimorphic, with a clear bimodal score distribution for all collections. The differences between Dart and Spitalfields are due to Spitalfields' wider spread of scores. This broader range indicates that Spitalfields males are more highly variable in the morphological expression of the subpubic concavity. The lack of intermediate scores for Dart is also interesting, and cannot be readily explained. In general, the scores for subpubic concavity are far less variable than for any of the skull traits used to determine sex.

5.3.9 Ventral Arc (Figures 4.9a to 4.9f, Chapter 4)

For the ventral arc, like the subpubic concavity, score distributions are all clearly bimodal and highly sexually dimorphic. The morphological expression of the ventral arc for Dart and Pretoria is less variable compared to that of the other collections, with fewer intermediate scores. Spitalfields and Coimbra females are slightly more variable in the morphological expression of the ventral arc compared to females of the other collections.

5.4 Overall Sex Determination Results

5.4.1 Pelvis

As noted in Chapter 4, no age-related differences in the percentages of correctly-sexed individuals using the pelvis alone were found, contrary to the findings of Walker (1995, 2005). Some differences were found, however, by sex: that is, for some collections, the differences between the proportions of correctly sexed males and females were greater than for others. For Grant, Coimbra, and Pretoria, the percentages of correctly identified males and females were quite close – the difference for Grant was 2.5%, for Coimbra, 0.3%, and for Pretoria, 0.1%. The larger Grant difference should be interpreted with some caution, again due to the low female sample size: of the 18 Grant females, only one was incorrectly sexed. A larger female sample may exacerbate or decrease the sex differences in correct identification using the pelvis. The differences in the proportion of correctly identified males and females for Spitalfields, Lisbon, and Dart are larger. For Spitalfields, females were sexed correctly more often, with a 5.6% difference between males and females. For Lisbon and Dart, males were sexed correctly more often than females, with a 3.7% and 5.8% difference, respectively. The results here do not agree with those of Meindl et al. (1985b: 84), where it was found that males were more often misclassified than females using the pelvis.

These differences suggest variation in the scale of sexual dimorphism between Dart, Lisbon and Spitalfields, particularly compared to Grant, Coimbra and Pretoria. For Coimbra and Pretoria, the pelvis performs equally well for males and females, suggesting that individual variation is the reason for the incorrectly-sexed individuals from these collections. Using pelvic morphology, the percentages of correct sex identification were closer in value across all samples for males than for females, indicating that females have somewhat greater variability in expression of pelvic morphology than males. Also interesting is that while pelvic morphology was most accurate for estimating the sex of Spitalfields females, the Spitalfields males fared the worst of the samples collected, while the opposite is true for the Dart sample. This suggests that these two collections have different ranges of expression of pelvic sexual dimorphism. The Dart sample has a scale of sexual dimorphism in terms of pelvic morphology that is skewed towards the "male", where higher numbers of females have narrower, more “male” pelves. Sexual dimorphism for Spitalfields’ pelves is skewed towards the “female” end of the scale (or scoring system), where higher numbers of males have wider, more “female” pelves. The total percentages for correctly sexed Lisbon males and female were not far from those of Dart, indicating that perhaps for Lisbon as well, the scale leans slightly more heavily towards the “male” side of the spectrum.

The overall high rate of correct sex determination, ranging from 94.8% to 97.3% for the sexes pooled, is further support for the reliability of the pelvis as a morphological sex indicator (also see Derry, 1909: 266; MacLaughlin and Bruce, 1990: 1384; Weiss, 1972: 239; Walker, 2005: 385; MacLaughlin and Bruce, 1986: 1380; Meindl et al., 1985b: 85; Bruzek and Murail, 2006: 227). In particular, the ventral arc and subpubic concavity, followed by the ischiopubic ramus ridge, are useful in the case of fragmentary pelvic remains.

5.4.2 Skull

While no age-related differences were found in terms of the percentages of correct sex determination using the skull alone, a clear sex-related trend was found in the collections under study: for all collections except Dart, females were more often sexed correctly than males. The small female sample size for Grant may again be reason for caution in interpretation of the sex differential; here, only a 1.1% difference was seen with a male bias. Sex differences for the other collections are substantially larger, particularly for the South African collections. Coimbra females are sexed correctly 8.4% more often than males; for Lisbon, this value is 10.7%. The values increase for the other collections: Spitalfields, 15.6%; Pretoria, 21.5%; Dart, 43.2%. These data support those of Meindl et al. (1985b: 84), where males were more often sexed incorrectly than females using the skull.

All of these collections (with the aforementioned exception of Grant) exhibit more “female” scaled skull traits compared to the skeletons used to develop the scoring method. The skull traits for males observed here were generally less robust, with less sexual dimorphism, than found in the reference collection used to develop the method (Walker, 2008: 40, used the Hamann-Todd, Terry and St. Bride’s collections; the method as originally developed used European individuals: Acsádi and Nemeskéri, 1970: 89). Furthermore, the Pretoria and Dart skulls were more “female” than those of Coimbra, Lisbon, and Spitalfields, as evidenced in the low proportions of males correctly identified using morphological traits of the skull. There also seems to be some geographical clustering in results. While the differences in percentages of correct sex determinations using the skull between collections were not very different, within-country results were more similar than results between other collections. That is, total percentages for Coimbra and Lisbon (both in Portugal) were close, as were total percentages for Dart and Pretoria (both in

South Africa). This was true of the pooled results and the males-only results, suggesting within-country similarities in skull morphology and scale of sexual dimorphism. Using the skull only for sex determination, Dart again had amongst the highest percentages of correctly-identified females, and the lowest percentage of correctly-identified males. This suggests that, compared to the reference collections used to develop the scoring criteria for the skull method, the scale of morphological skull sexual dimorphism for Dart is skewed towards the "female" end of the distribution. The Pretoria results are also interesting. The skull performed least successfully for Pretoria females, and second-lowest in terms of allocation accuracy for Pretoria males. This suggests that the Pretoria male and female skull morphology may exhibit more overlap (less sexual dimorphism) than was recorded in the other collections.

The skull did not perform as well as the pelvis did as a single element, but this was expected. As noted above, and in Chapter 3, the pelvis is generally held to be the most reliable morphological sex indicator due to its functional role in females for childbirth. In all collections for the pooled results, the skull alone did not perform as well as the pelvis alone. For the females, the skull alone did not perform as well as the pelvis for all collections except Dart; for Dart females, the skull performed marginally better (92.0%) compared to the pelvis (91.8%). In absolute terms, these differences are the result of five of 73 incorrectly sexed females using the pelvis, and of six of 75 incorrectly sexed females using the skull. For the males, the pelvis performed better for all collections by itself than the skull alone. When the sexes were pooled, the pelvis performed better for all collections.

5.4.3 Pelvis and Skull Combined

Again, no age-related trends were observed when the skull and pelvis were used together for sex determination. Here, no clear differences by sex were observed – for Spitalfields, Coimbra, and Pretoria, females were more often sexed correctly than males, while the reverse occurred for Grant, Lisbon, and Dart. Meindl et al.'s (1985b: 84) conclusions, that males are more often misclassified than females, are thus not supported particularly well here. The reverse, that females are more often misclassified than males (e.g. Mays and Cox, 2000: 125), is equally not well supported here. This lack of conclusive direction in misclassification likely reflects the variation in the range of dimorphism expressed in the collections studied by these other authors.

Taken by itself, the pelvis performed better and is a more reliable sex indicator than the skull, thus supporting data from other studies (Derry, 1909: 266; MacLaughlin and Bruce, 1990: 1384; Weiss, 1972: 239; Walker, 2005: 385; MacLaughlin and Bruce, 1986: 1380; Meindl et al., 1985b: 85; Bruzek and Murail, 2006: 227), as noted above. The suggestion that using multiple indicators is best (e.g. Meindl et al., 1985b: 85) is supported by the results showing that the pelvis and skull, when used together, provide more accurate results than does the skull alone. However, the same cannot be said for the pelvis and skull combined results compared to the results for the pelvis alone. For Coimbra, Dart and Pretoria males and for Lisbon males, females and the sexes pooled, the results for the pelvis alone compared to the pelvis and skull were exactly the same. Perhaps for Lisbon, and for Coimbra, Dart and Pretoria males, when pelvis and skull data did not agree on sex, the evidence from the pelvis was clearly indicative of sex, while the skull was "undetermined", thus leaving the pelvis as the deciding sex indicator, even when both skull and pelvis were present. For Grant, the female disparity is simply due to differences in absolute numbers of females with skulls and pelves available for study; there were 18 pelves and 13 skulls available for data collection, and only one individual incorrectly sexed using each element (but for different individuals; when pelvis and skull data were combined to determine sex and did not agree, the pelvis data were weighted more heavily and provided the deciding factor). The reason for the percentage difference was simply that one individual of 13 represents a higher percentage than does one individual of 18. The same is true for males alone and pooled results for Grant. The pelvis and skull combined resulted in the same absolute number of incorrectly-sexed individuals than occurred using the pelvis alone, but fewer individuals had both the skull and pelvis available, and those sexed incorrectly accounted for a higher percentage each. The reasons for the lower percentages of correct sex determination for Spitalfields females and sexes pooled when using the pelvis and skull compared to pelvis alone are the same. For the other categories, that the percentage of correct sex determinations using the pelvis alone and the pelvis and skull together remains the same is not surprising, given that the pelvis was weighted more heavily than the skull when there were disagreements in sex assessment (as is standard practice).

5.4.4 Metrical Data

Generally, the results show that modification 1 is best for determining sex compared to the other modifications available. However, when skeletal elements or landmarks necessary for the measurements are damaged or missing, the other modifications generally perform fairly, but variably, well. The allocation accuracies for modification 1 varied by collection, from 100% for the Grant Collection (sexes pooled) to 89.6% for the Pretoria Collection (sexes pooled). The other modifications performed most poorly for the Dart Collection, with 76.9% allocation accuracy for the sexes pooled, and best for Grant and Pretoria, with 100% allocation accuracy each for the sexes pooled. Generally, this supports Albanese's (2003a: 7) suggestion that modification 1 produces the best results, and should be used where possible (i.e. where all skeletal elements necessary are present, the best allocation accuracy results from modification 1, the equation using hipbone height, iliac breadth, SPRL, maximum diameter of the femoral head, and epicondylar breadth). However, the results here did not match Albanese's allocation accuracy of

98.5% (on a target sample from one of the same collections used as the reference sample in his study), providing further evidence of variation in sexual dimorphism.

Comparison of the amount of sexual dimorphism for each measurement is also interesting; the clustering of values for Coimbra and Lisbon and Dart and Pretoria for most measurements suggests that there are more similarities within-country than between, for the measurements used. However, for iliac breadth and SPRL, the Coimbra/Lisbon cluster does not appear, suggesting that sexual dimorphism is variable by skeletal element. Grant and Spitalfields, with their differing geographic origins, follow more independent paths of sexual dimorphism, with values varying by skeletal element, not clustering with any other sample in particular. However, it must be noted that Grant individuals are largely of European origin (Bedford et al., 1993: 288). Overall, the evidence seems to support that within-country similarities tend to be present, supporting the extrapolation of methods developed using a particular reference collection to skeletons of similar geographic origin.

Also of interest is that within-country proportions of correct sex prediction are more similar to each other than to that of the other collections. Using modification 1, Coimbra and Lisbon females, for instance, both had percentages of correct sex identification of 100.0%, while for males the values were 84.1% and 88.9%, respectively. Meanwhile, the values for Dart and Pretoria were also quite similar – 87.5% of Dart females and 89.1% of Pretoria females were sexed correctly as were 98.6% of Dart males and 90.1% of Pretoria males. Spitalfields females were sexed correctly in 96.3% of cases, as were 95.5% of Spitalfields males; Grant males and females were sexed correctly in 100.0% of cases.

The amount of disparity in sex estimation using metrical analysis for Coimbra, Lisbon and Dart is problematic; ideally, sex determination methods should work at least nearly equally well for both sexes (Albanese, 2003a: 8). Further, the results here do not agree with Albanese's (2003a: 7) results of 98.5% allocation accuracy for males and females separately, despite the fact that part of Albanese's target test sample was also drawn from the Coimbra Collection (the other part from the Terry Collection). If the disparity in allocation accuracy between the sexes is great, it suggests that sexual dimorphism of that particular skeletal feature is low for the target population, or that the target population's scale of sexual dimorphism is shifted compared to that of the reference collection. The criterion for modification inclusion given by Albanese (2003a: 8) was that the difference between the sexes in allocation accuracy had to be less than 5%. While modification 1 seems fairly successful in determining sex for Coimbra, Lisbon and Dart (92.9%, 94.6%, and 93.5%, respectively), this overall success is negated by the fact that for Coimbra and Lisbon males and Dart females, the success rate for each of those groups is less than 90% (the other reliability criterion for inclusion of modifications given by Albanese, 2003a: 8). The sex differences in allocation accuracy for Dart, Lisbon and Coimbra do not adhere to the 5% criterion. If held to the standards of the expected allocation accuracy of the innominate bone using morphological methods given by some forensic papers of 90 to 95% accuracy (e.g. MacLaughlin and Bruce, 1986: 1380), that Albanese's method performs less well for the Lisbon and Coimbra males and Dart females is also problematic. However, the expected and accepted standards of allocation accuracy in archaeological contexts may be lower (Scheuer, 2002: 299). Furthermore, where stated accepted levels of allocation accuracy are published, they appear to refer to the pooled results, not the sexes individually; as such, Albanese's method does perform to the forensic standard when the results are viewed overall, for all modifications combined. It is also useful for fragmentary skeletal material.

That Albanese's (2003a) method did not perform particularly well for Coimbra males is interesting, as the Coimbra Collection was used in the method's development, alongside the Terry Collection. Albanese (2003a: 3) does note that in choosing the reference sample, an upper age limit of 79 was arbitrarily used to diminish the effects of pathological joints (probably largely osteoarthritis) or decrease in skeletal robusticity with age that might result in misleading or missing data. However, despite Coimbra males with ages-at-death in the 80s and one aged between 90 and 99, there does not seem to be an age-related decrease in allocation accuracy. Instead, incorrect sex determinations are found in all age groups except 20 to 29 and 90 to 99 years. Thus, the upper age limit of the reference collection has not impacted the allocation accuracy of the method as applied to Coimbra males as a target sample. Perhaps, if Terry males have relatively larger measurements than Coimbra males, the inclusion of the Terry sample with Coimbra individuals in the method's development set the method's scale of sexual dimorphism to slightly more "male" than needed for Coimbra alone (as a target sample).

5.5 Variability in Sexual Dimorphism

In terms of cranial morphology, the least sexually dimorphic single indicators were the mastoid process and supraorbital margin. Similarly, Williams and Rogers (2006: 733) noted that the supraorbital margin did not perform well on their study of Americans of European white ancestry from the Bass Collection. The glabella was found to be very useful in determining sex in the historical St. Thomas Anglican Church cemetery sample, from Belleville, Ontario, where individuals were of European (mostly British) origin (Rogers, 2005: 496, 499); nuchal crest and mental eminence were deemed to be of 'secondary value', while the mastoid process was of 'tertiary consideration' for that sample. Interestingly, the glabella displayed the least sexual dimorphism in the South African collections, but was the most sexually dimorphic single indicator for Spitalfields and Coimbra, highlighting the variation between populations from different geographic locations, and the possibility of shared characteristics between geographically close populations. Population differences were found in Walker's (2008: 48) study using the Hamann-Todd Collection, Terry Collection, St. Bride's cemetery sample and an archaeological Native American sample (California, USA). The modern Americans (black/African American and white/European American) were found to be more robust than the English sample, with higher average scores. Within-population difference was also found. The European American males had more robust glabellas and more rounded supraorbital margins than the African American males who had, in turn, more pronounced mental eminences than the European American males. In general, the archaeological Native Americans were less sexually dimorphic and more robust than the modern population samples (Walker, 2008: 48). While the population differences are supported by the evidence presented here, the glabella differences are most interesting in this context, as the African Americans had less robust glabellas, as did the (South) Africans studied here. Perhaps some geographical- or ethnicity-related differences in sexually dimorphic traits remain despite the effects of immigration and generational differences.

Despite other studies finding contrary results (Meindl et al., 1985b: 81; Walker, 1995: 37, 40; Walker, 2005: 385), no age-related trends were found in any of the collections under study here – that is, older females and younger males seemed no more likely to be misclassified than younger females and older males. Similarly, Mays and Cox (2000: 126) note that Molleson and Cox (1993) found no increase in craniofacial robusticity in the older females at Spitalfields. Other studies on other documented collections have similarly not found any age-related trends in misclassification of sex (Williams and Rogers, 2006: 732).

While sexual dimorphism of most of the cranial indicators used here can largely be attributed to sexual dimorphism of size and robusticity, craniofacial robusticity is also affected by diet (Mays and Cox, 2000: 125). Processing tough foods leads to an increase in masticatory muscle size and associated cranial features, so for populations where the diet is heavy in coarser foods, craniofacial features may seem more "male" regardless of biological sex. Furthermore, as Mays and Cox (2000: 125) point out, archaeological populations tended to have diets with tougher foods than do modern populations, which are used as reference collections for the development of sex determination methods. Conversely, the modern reduction of coarse, tough foods in the diet is suggested to have lead to a decrease in masticatory stress and resulting changes in (diminishing) craniofacial robusticity (Jantz and Meadows Jantz, 2000: 335). The use of the jaws and teeth as tools, or other cultural practices, may also lead to increased robusticity of cranial features (Mays and Cox, 2000: 125). Other studies have implicated environmental factors and improvements in nutrition and medical care as the driving forces for secular changes in craniofacial morphology, although the effect on sexually dimorphic cranial features is not mentioned (Jantz and Meadows Jantz, 2000: 335; Buretić-Tomljanović et al., 2006: 674; Gonzalez et al., 2010: 377). Genetic factors are also a possible reason for craniofacial change (Jantz and Meadows Jantz, 2000: 335-336); still others suggest that demographic changes and shifts towards earlier maturation, including the modern decrease in infant and child mortality, decreased fertility rates and increased longevity, are possible reasons (Weisensee and Jantz, 2011: 556-557). Some of the skeletal measures studied by Weisensee and Jantz (2011) in their examination of secular change in cranial morphology incorporated areas used for sex estimation, including the mastoid process and glabella, although the impact of secular change in terms of sexual dimorphism was not part of that study.

The significant differences found between various collections serve to highlight the variability in sexual dimorphism across populations. Furthermore, the skeletal traits scored to determine sex do not necessarily vary in the same direction, even when the traits are from the same skeletal element. For example, of the skull traits scored, the supraorbital margin and nuchal crest score distributions were found to vary significantly between the Lisbon and Coimbra Collections – these Portuguese collections represent people from cities quite close geographically, only about 210 km apart. This supports the body of evidence suggesting that such differences can occur in populations within close geographic proximity to each other; Washburn (1949: 428), for instance, noted that South African ‘Bushmen’ had wider sciatic notches than did South Africans from the Bantu tribes (although it was noted that the Bushmen skeletons were curated from a wider geographic range).

Further geographic clustering was found in terms of the percentages of sexual dimorphism for some of the measurements taken for Albanese’s (2003a) metrical method: the maximum length, maximum diameter and epicondylar breadth of the femur, hip bone height and AIL all show clustering between Lisbon and Coimbra (Portugal), and Dart and Pretoria (South Africa). This suggests that with regard to sexual dimorphism in these femoral and pelvic measurements, within-country similarities are found. However, while the percentage of sexual dimorphism for the SPRL measurements cluster for Dart and Pretoria, they do not for Lisbon and Coimbra – not every pelvic dimension, then, shows such within-country similarity in the scale of sexual dimorphism.

Using skull traits alone, Williams and Rogers (2006: 732) found a 96% allocation accuracy, with higher male accuracy than female, and Rogers (2005: 497) found an allocation accuracy of 89.1%; however, 17 morphological traits were examined in their study compared to five used here. Walker’s (2008) version of the method was tested on a modern forensic Balkan sample from mass graves near Belgrade; sex was identified via soft tissue and ‘obvious sexual characteristics’ at the University of Belgrade (Đurić et al., 2005). Đurić et al. (2005: 160-161, 163) found a fairly low allocation accuracy of 70.56%, with more males incorrectly labelled female; however, the scoring system was simplified to 1 to 3, which may have affected accuracy. Here, only the Grant Collection sex determinations using the skull produced similarly high allocation accuracy to the studies listed above, at 93.2% for the sexes pooled (with similar male and female results). The skull was not as successful for the other collections (ranging from 60.0% for Pretoria to 87.0% for Lisbon), and female allocation accuracy was higher than that of males for all collections except Grant, contrary to the findings of Williams and Rogers (2006: 732). The temporally later European or European-origin collections had allocation accuracies closer to those found by Rogers (2005: 497), at 85.0% for Coimbra and 87.0% for Lisbon, as well as the aforementioned Grant results. This reflects the fact that the original morphological sex determination method using the skull seems to have been developed or at least tested on European skeletons (Acsádi and Nemeskéri, 1970: 89). This provides further evidence for regional and perhaps temporal variation in sexual dimorphism. Acsádi and Nemeskéri (1970: 91) also observe a 'gracilization' in sexual dimorphism from Upper Paleolithic skeletons onwards, suggesting that the use of morphological methods of sex determination on populations other than the 'Europoid main race' should be done after taking into consideration the degree of sexual dimorphism of the target population. Rogers (2005: 497) notes: 'Hrdlicka stated that experienced investigators should be able to correctly identify the sex of an unknown skull in 90% of cases. Stewart successfully determined the sex of 100 crania from the Terry skeletal collection with 77% accuracy. Krogman and İşcan optimistically concluded that 92% accuracy could be achieved.' However, experience of the investigator aside, it is rather the differences in populational sexual dimorphism that are reflected in differences in allocation accuracy; Scheuer (2002: 299) similarly notes that while some researchers claim up to 98% accuracy using pelvis indicators, differences in sexual dimorphism between the target and reference collection may give poorer results.

Generally, the pelvis is considered a more reliable sex indicator than the skull, and this was supported by the results presented here. In terms of individual skeletal elements used for determining sex, the sciatic notch was the least sexually dimorphic of the single pelvic indicators, with the lowest percentages of correct sex identification for both males and females for all collections. The most sexually dimorphic single pelvic indicators are the subpubic concavity and the ventral arc.

Rogers and Saunders (1994: 1048-1049) tested 17 pelvic traits on the St. Thomas Church cemetery site, from Belleville, Canada. The allocation accuracies for the traits relevant to this study are as follows: ventral arc: 86.9%; sciatic notch: 85.7%; subpubic concavity: 83.8%; ischiopubic ramus ridge: 80.0% (Rogers and Saunders, 1994: 1051). The sciatic notch allocation accuracies in this study are all lower than that reported by Rogers and Saunders. The closest value to the 85.7% found by Rogers and Saunders was the 81.3% correct sex determination for

Coimbra males. The allocation accuracies for the ventral arc and subpubic concavity are higher for all collections studied here than for the St. Thomas sample. The ventral arc allocation accuracies are approximately 5% to 10% higher and subpubic concavity allocation accuracies are approximately 10% to 14% higher compared to the results reported by Rogers and Saunders. For the ventral arc, only Grant had a sex difference of more than 5%, and while Pretoria and Spitalfields had sex differences in allocation accuracy for the subpubic concavity of over 5%, the allocation accuracies for males and females in both collections was over 90% each. Allocation accuracies for the ischiopubic ramus ridge in this study are comparable to that found by Rogers and Saunders. While it is difficult to compare the allocation accuracies of this study to those of Rogers and Saunders (1994) due to possible interobserver differences, that the ischiopubic ramus ridge values are comparable lend some credence to considering the differences in the other pelvic traits as further evidence for interpopulation differences in sexual dimorphism.

Of the pelvic indicators used in the Phenice method, the sciatic notch has undergone testing by a number of researchers examining its efficacy as a sex identifier and whether its morphology and scale of sexual dimorphism varies by population. In terms of population variation, Walker (2005: 388) found that American samples did not significantly differ from one another, but both differed significantly from the English sample. The St. Bride's individuals were found to have a more feminine morphology than the Americans – English females scored 1 more often, and English males score 1 or 2 more often than the Americans (Walker, 2005: 388). Here, while no differences for female sciatic notch distribution were significant, Spitalfields (English) females also scored 1 more frequently than did females from the other collections. However, Spitalfields males did not seem to have lower scores more often than did other collections. Spitalfields males did have a flatter score distribution, and were less highly peaked at the definitively "male" scores of 4 or 5 than did males of other collections, perhaps indicating a slightly more variable sciatic notch morphology. For MacLauglin and Bruce (1986: 1384, 1389), the sciatic notch performed well in discriminating English females, but poorly for Dutch females. This difference was attributed to a possible difference in the allometric relationship between sciatic notch width and overall size of the body (the Dutch females were the largest in body size out of their samples). In this study, the sciatic notch performed variably (but generally poorly) for females and males – it worked best for Grant and Spitalfields females (European origin and English) and worst for Coimbra females. The sciatic notch performed better for males in all collections, best for Coimbra males and worst for Lisbon males. Coimbra's scores for the sciatic notch are skewed towards the male morphology, and the poor performance for females and good performance for males suggests low sexual dimorphism of the sciatic notch for this collection. Some age differences were also reported – sciatic notches tended to decrease in width (become more masculine) with increasing age, although this effect diminishes after age 50 (Walker, 2005: 389). Here, no age-related trends were seen.

The Phenice method was developed on 275 individuals from the Terry Collection (Phenice, 1969: 298, 300), and Phenice listed an allocation accuracy of ‘at least 96%’. However, other researchers have not been as successful with the Phenice method. Bruzek (2002: 157) suggests that inconsistencies in the allocation accuracy of the Phenice method are a result of observation of the pubis alone, and not the entire innominate bone, but estimates (based on a number of other studies) that allocation accuracy is likely around 80% for the method. Accuracy in Lovell’s (1989: 117, 120) study on 36 dissecting-room known sex individuals (curated at Simon Fraser University, British Columbia, Canada) was lower than in Phenice’s study. The highest accuracy in her study, by the most experienced observer, was 92%, while the average accuracy of her less experienced observers was 83% ± 7% (Lovell, 1989: 118). Lovell (1989: 199) suggested that as the age distribution of her sample was older than that of Phenice, it was possible that age affected expression in the tested traits, resulting in decreased accuracy. In their study of three known sex European samples, including a sample from St. Bride’s Church (17th - 18th C, London, England), and two modern dissecting-room samples (from Scotland, curated at the University of Aberdeen, and the Netherlands, curated at the University of Leiden), MacLaughlin and Bruce (1990: 1387) found accuracy to be highest in the English collection, at 83%, while only 68% of the Dutch collection and 59% of the Scottish collection were accurately sexed. The subpubic concavity was found to be the most reliable single indicator out of the three features tested. However, for sex to be assigned in this study, two out of the three traits had to correlate, while Phenice’s method was less stringent (MacLaughlin and Bruce, 1990: 1390). Rogers and Saunders (1994: 1054), using the Phenice combination of traits (out of 17 pelvic traits they tested) on 49 adult Canadian (European origin) skeletons from the 19th century St. Thomas Church cemetery, found an allocation accuracy of 88%. Other tests of the Phenice method alone found a 1.2% error rate in sex determination of male pelves, and a 1.9% error rate for females (Konigsberg et al., 2002; Steadman et al., 2006). In this study, the allocation accuracies for the sexes pooled were all fairly high, ranging from a low of 94.8% for the Dart Collection to 97.3% for the Pretoria Collection. These values are closest to those of Phenice, despite different geographic origins of the collections compared to the US origins of individuals in the Terry Collection. Temporally, however, the Terry Collection is most similar to the Grant, Lisbon and Coimbra Collections (Spitalfields is composed of earlier dated individuals, while Dart and Pretoria are composed of more recently dated individuals). The results here are better than Lovell’s; while she suggested that an older age distribution may be the reason for lower allocation accuracies, the sample used here all include old-age individuals (as did her study). While interobserver differences cannot be discounted, it is possible that population variation is the reason for allocation accuracy differences between Lovell’s study and this study.

Also interesting is that the Spitalfields allocation accuracy is higher than that reported for the St. Bride's church cemetery sample (MacLaughlin and Bruce, 1990: 1387), despite very similar geographic origins (both church cemeteries were located in London, UK, although the Spitalfields individuals are largely of French Huguenot descent), social status (e.g. Molleson and Cox, 1993; Scheuer and Bowman, 1995; Megyesi et al., 2006: 364), and overlapping time periods. While interobserver differences and population differences in sexual dimorphism expression are possible reasons, the methodological differences in MacLaughlin and Bruce's study, mentioned above, may also be a reason for wide differences in allocation accuracy between Spitalfields and St. Bride's.

Most studies of sex determination method efficacy have examined only cranial or pelvic traits, but Đurić et al. (2005: 161, 163) found that when skull traits were used in combination with visual assessment of pelvic traits, allocation accuracy was much higher (from 95 to 100%). However, a total of 16 traits were assessed in that study, a larger suite of traits than used here. In any case, while allocation accuracy of the pelvis and skull combined in this study did not reach 100%, the combination provided the most consistently high accuracy, ranging from at 95% to 97.5%, within the range reported by Đurić et al. (2005: 161, 163). The absolute highest accuracy was for Albanese's metric method; accuracy was 100% for the Grant Collection only.

These data clearly have implications for the application of the methods to skeletons of unknown provenance, either archaeological or forensic. While Scheuer (2002: 299) noted that 85% allocation accuracy seems to be sufficient for archaeology, higher accuracy is necessary in forensic contexts. Bruzek and Murail (2006: 226) list 95% allocation accuracy as acceptable for forensic purposes; Rogers (2005: 499) does not list a specific level of accuracy (for the USA and Canada), but rather that admissibility in court requires knowledge of potential or known error rates (among other criteria). Certainly, the use of all possible skeletal traits, or at least the ones tested in this study should be used whenever possible.

5.6 Age Estimation Methods

The results presented in the previous chapter provide clear evidence of population variation in ageing rates for the age estimation methods tested – the significant differences between various collections found in mean ages per phase or score are the evidence for that phenomenon. Significant differences in the shape of phase (or score) distributions were also found between some collections, despite similar age distributions of the samples. This provides further evidence for variability in ageing, but adds a different dimension to such variability. Significant differences in phase distribution suggest that a particular skeletal characteristic(s) deemed important for placement into a particular phase for the method under consideration does not occur at the same frequency or with the same timing in some populations compared to the reference collection (in relation to other characteristics of the same phase). Here, investigation of the skeletal characteristics scored individually using the Buckberry-Chamberlain auricular surface method enabled deeper investigation into the specific age-related characteristics that differ between populations.

First, the rough relationship of phase with age for the ageing methods must be briefly discussed because if there is no relationship of increasing phase numbers with increasing age, then any discussion of variation in mean age per phase is not necessarily indicative of variability in ageing rate. Mean ages were calculated using Microsoft Excel, and tables with mean ages per phase, standard deviations and age ranges are given by ageing method, for each collection in Tables A11.1 to A11.6, Appendix 11.

For the Suchey-Brooks pubic symphysis method, there is a relationship between phase and age, with mean age generally increasing by at least 10 years between subsequent phases, with the exception of the mean age difference between phases I and II. However, had individuals younger than age 20 been included, this problem may have been alleviated. The age ranges for Phases III, IV, and V tend to be very wide.

The Meindl-Lovejoy auricular surface method also shows a relationship between phase and age, but with smaller mean age increases between subsequent phases compared to Suchey-Brooks – however, there are more possible phases used in the Meindl-Lovejoy method compared to that of Suchey-Brooks, so this is not unexpected. More problematic are the stochastic variations in mean age per phase that appear in a few cases – for Lisbon, the mean age for the 45-49 year phase is two years higher than for the subsequent 50-60 year phase, as is the 40-44 year phase for Dart compared to the subsequent 45 to 49 year phase. Age ranges are again very wide for the middle phases.

The Buckberry-Chamberlain auricular surface method seems to show a slightly better relationship between phase and age compared to the Meindl-Lovejoy method, in that mean age consistently increases with phase, for all phases, across all collections. Age ranges are again very wide for the middle phases.

While lateral-anterior cranial suture closure does show a relationship with age, in many cases, only a very small increase can be seen in mean age for subsequent phases. Accompanying this are extremely wide age ranges. Vault cranial suture closure also shows a relationship with age, and with slightly more widely-spaced mean ages for subsequent phases compared to lateral-anterior sutures (although the difference is still only one to three years for some cases). Age ranges, however, are still extremely wide. Figure 5.3 illustrates the lateral-anterior suture closure age range for Coimbra – scores of 7, for example, occur in every known age group. For the fourth rib method, mean age does not consistently rise with score at the higher scores, and there are only very small increases in mean age for the lower scores; this is clearly not beneficial for an age estimation method. However, small sample sizes may be part of the problem for the fourth rib in this study.

As all the age estimation methods do show a relationship between phase (or score) and age, discussion can continue regarding variability in ageing rates between populations.

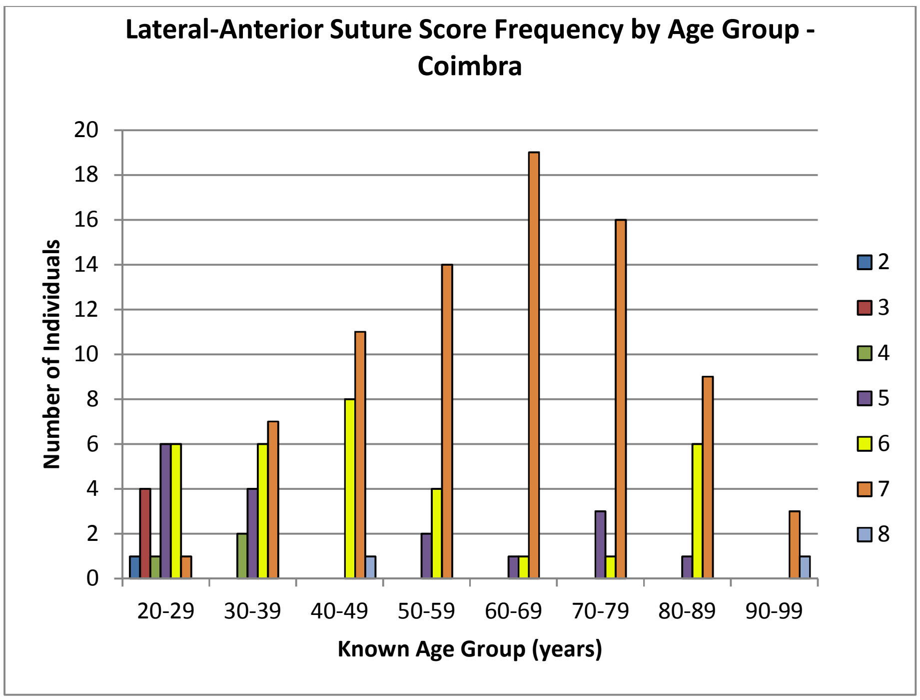


Figure 5.3. Bar chart of lateral-anterior suture score frequency by age group for Coimbra

5.7 Sex Differences in Rates of Ageing – Within Collection

While some significant sex differences in mean age were found for the Suchey-Brooks pubic symphysis and sternal end of the fourth rib methods, it is not particularly concerning, as separate male and female standards are already used. The differences in the other ageing methods tested, however, are potentially more revealing, as there are no sex-specific standards for these.

The Buckberry-Chamberlain method shows significant differences in variance in phase II between Coimbra and Pretoria males and females, but this seems to be the result of just one or two slightly older females in this phase. Phase VII's significant differences in variance for Coimbra and Dart do not reveal any trends. Coimbra's males had a wider variance and were clustered at younger ages than were the females, while Dart's females had the wider variance, with more individuals being younger than the males. For phase III, Coimbra females had a significantly lower mean age than did the males. Most females in the 20 to 29 year age group exhibited an auricular surface morphology placing them in phase III, while males aged 20 to 29 were more likely to be in phase II (although two were in phase IV, and one in phase V). For Coimbra, then, young females show a slightly older-looking auricular surface morphology than do young males.

The Meindl-Lovejoy auricular surface method has more possible age categories (eight) compared to the Buckberry-Chamberlain method (seven). Interestingly, the only significant differences found between the sexes using the Meindl-Lovejoy method do not match those found using the Buckberry-Chamberlain method. Only Pretoria males and females in the 40-44 year and 45-49 year categories show significant differences in mean age; no other significant differences were found. This serves to highlight the differences in skeletal characteristics and qualities emphasised in each of these auricular surface methods, perhaps as a result of the different reference collections used in each method's development. The Pretoria differences in the 40-44 year and 45-49 year categories are not in the same direction; females have the higher mean age in the 40-44 year category, while for the 45-49 year category, males have the higher mean age. As these trends are not consistent, and do not occur in all Meindl-Lovejoy categories, it is difficult to suggest whether sex-specific standards are necessary or not. Stochastic variation is still a possibility, so it seems premature to suggest sex-specific standards based on two significantly different values.

Similarly, the few significant differences in mean and/or variance for cranial suture closure phases and sternal end of the fourth rib phases are not consistent and do not show any particular trends. Small sample sizes for the sternal rib end method further complicate the picture. Results from this study do not seem sufficiently robust to suggest that sex-specific standards are necessary. However, when boxplots of mean ages for Meindl-Lovejoy phases by sex are examined for each collection (Figures A12.1 to A12.6, Appendix 12), it becomes clear that female mean ages tend to be higher for nearly every phase, for every collection except for Coimbra. Differences are not significant, but do exist. The same is true for lateral-anterior cranial suture closure for Lisbon.

Significant differences in mean age were found between males and females only in certain phases, but not across all of any one method's phases. This suggests that the particular skeletal characteristic or quality that is definitive of placement in that phase is the reason for the sex differences. This is not true necessarily for methods where placement in a phase is the result of summing of scores for a number of characteristics of the particular skeletal element (or for cranial suture closure, by anatomical location), because varying combinations of scores could result in placement in the same phase.

5.7.1 Buckberry-Chamberlain Trait Scores

The mean and variance were also compared within collections by sex for each Buckberry-Chamberlain scored trait. As with the phases for each method, while some significant differences were found in either mean or variance between males and females of some collections for some traits, no consistent differences or trends were found. Again, the evidence seems insufficient to suggest separate male and female scoring standards. However, it is interesting that for transverse organisation, only Spitalfields had enough males and females with scores of 1 to calculate mean and variance, given that Spitalfields was used as the reference collection on which the method was developed. The sampling strategy was such that an attempt was made for equal numbers of individuals to be sampled for each age group, including the youngest groups (20 to 29 and 30 to 39 year olds), who presumably are the most likely target groups for scores of 1. Thus, that only the youngest Spitalfields individuals scored 1 for transverse organisation suggests a difference in ageing rate for this group, while the youngest individuals from other groups tended to score more highly. This is somewhat problematic as Spitalfields was a reference sample for developing the method, and this suggests that individuals for the same age group but from other samples with different ageing characteristics (like those from all of the other collections sampled here) have the potential to be slightly overaged using this method.

5.8 Phase/Score Distributions and Median Differences – Ageing Variability

5.8.1 Suchey-Brooks

That both Lisbon and Spitalfields had significantly different distributions of pubic symphysis phases to Grant is interesting in that the Grant Collection seems to have the “oldest” distribution of phases; the relative proportions of individuals in phases V and VI are higher than in any other collection. Meanwhile, the phase distributions of Lisbon and Spitalfields are younger than the other collections, with higher proportions of phases III and IV than the other collections. Bar charts of the phase distributions for Lisbon and Grant are shown below (Figures 5.4 and 5.5) to illustrate the differences in phase distribution. While only the differences between “youngest” and “oldest” distributions are significant, even this relatively good indicator of age does display variability in ageing rate between populations.

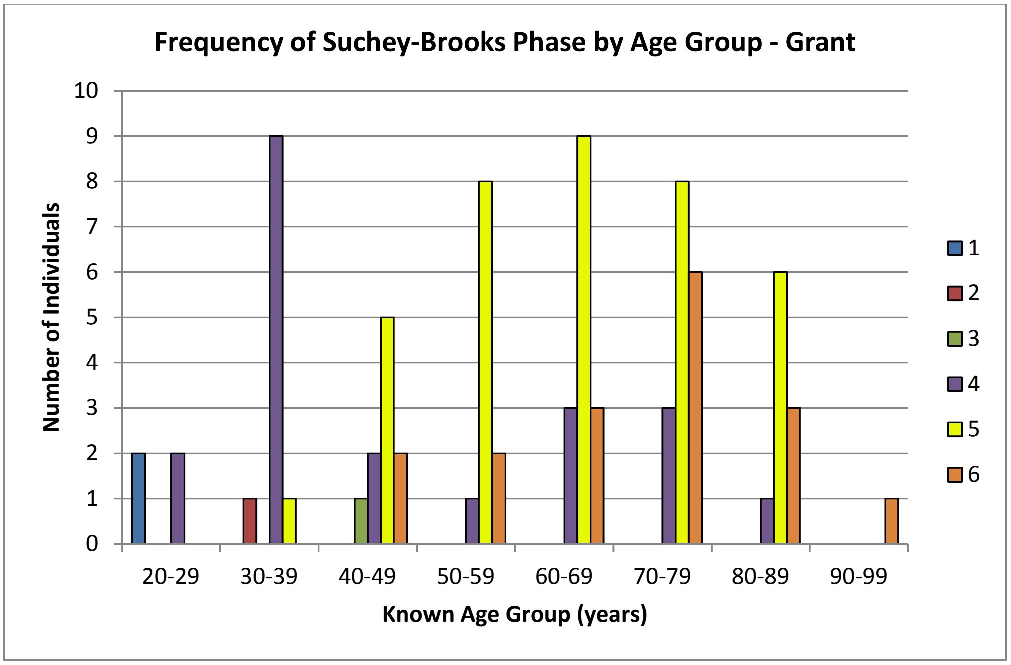


Figure 5.4. Frequency of Suchey-Brooks phase by known age group for the Grant Collection

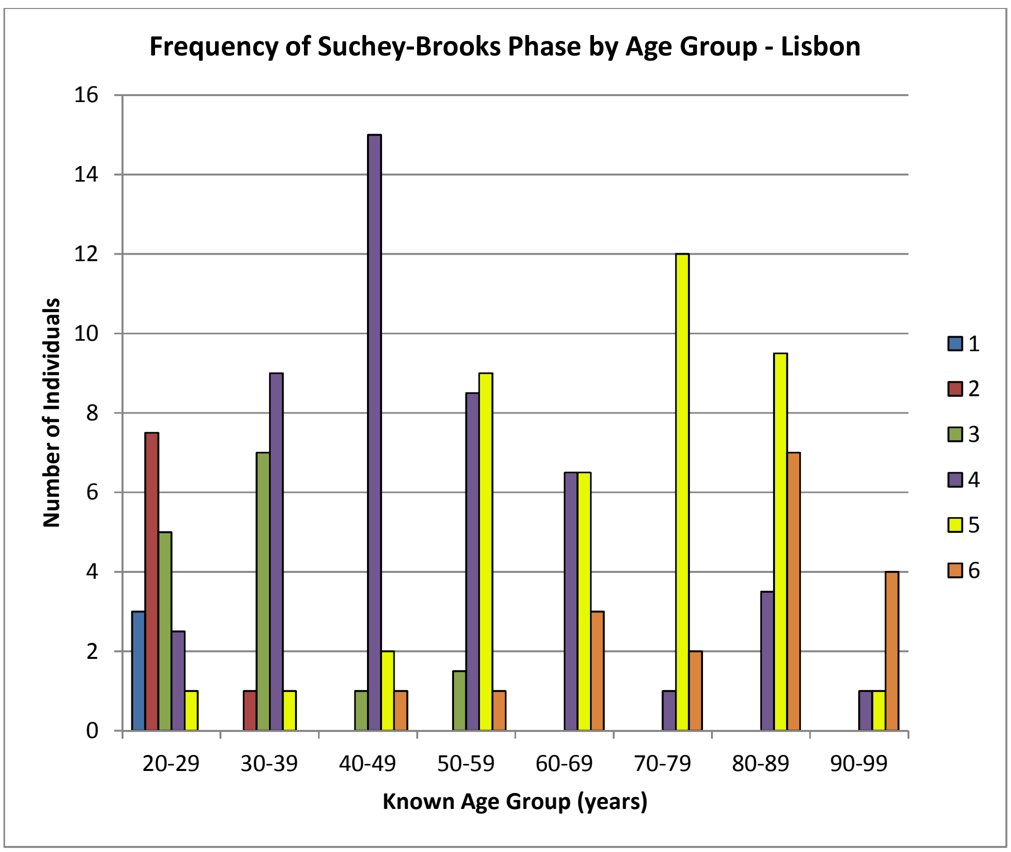


Figure 5.5. Frequency of Suchey-Brooks phase by known age group for the Lisbon Collection

5.8.2 Meindl-Lovejoy

It is interesting that the most common phase frequency for the Meindl-Lovejoy pubic symphyseal method is clustered by geographic location. The 30-34 year phase has the highest number of individuals for the Portuguese collections, while the 35-39 year phase has the highest number of individuals for the South African collections. Spitalfields and Grant also share a most-common phase, of 50-60 years. Few Portuguese or South Africans (particularly from the Pretoria Collection) were placed in the highest phase (60+), but this phase was more common for Spitalfields and Grant. Spitalfields and Grant lie on the “older” end of the ageing scale for the auricular surface, while the Portuguese collections display the youngest auricular surfaces, followed closely by the South African collections.

These clusters and the significant differences in phase distribution speak to variability in auricular surface ageing rates as well as the possibility of geographic proximity resulting in similarity in ageing rates.

5.8.3 Buckberry-Chamberlain

It is not surprising that the significant differences in the Buckberry-Chamberlain auricular surface results are between different combinations of collections compared to the significant differences found in the Meindl-Lovejoy results, despite both methods using the auricular surface. This is because the Buckberry-Chamberlain method relies on scoring individual characteristics of the auricular surface, then combining these scores to form a composite score, with an associated phase and age range.

The Meindl-Lovejoy method, on the other hand, places any particular auricular surface directly into an age range (by choosing the most appropriate suite of auricular surface characteristics). Thus, a number of Buckberry-Chamberlain score combinations, potentially reflecting very different suites of characteristics (although hopefully limited by each characteristic’s connection with ageing and degeneration), end up with the same composite score or phase, which is reflected in the variability of significant differences found in phase distribution and median.

Another indicator of differences in rates of ageing comes by way of proportions of individuals with the lowest and highest scores. Despite similar numbers of individuals in the 20 to 29 and 30 to 39 year age groups, Coimbra and Spitalfields had lower numbers of individuals in Meindl-Lovejoy’s first category (20-24 years) compared to Lisbon, Dart and Pretoria. Even if differences are not significant, they are still evidence for differences in ageing rate. For example, younger individuals from Coimbra and Spitalfields exhibit more advanced, older-looking auricular surface morphology than younger individuals from the other collections. For some methods, mean and variance could only be calculated for certain collections for the lowest or highest phases, as other collections did not have enough individuals in that phase for measures of central tendency and dispersion to be calculated. Perhaps for the oldest ages (and highest scores), differences in availability of the oldest old (those in their 80s and 90s) are the issue, but the youngest age categories were little affected by availability. Thus, where, for instance, only the Dart Collection had enough cranial vault scores of 2 to calculate mean and variance, a difference in ageing rate is suggested, as young individuals from other collections had consistently higher scores.

5.8.4 Lateral-Anterior Cranial Sutures

The significant differences in Spitalfields' phase distribution and the phase frequencies again suggest that Spitalfields exhibits a more advanced, older-looking morphology. Besides the more advanced rate of ageing in terms of lateral-anterior cranial sutures, the highly peaked distribution of phases also highlights the shape differences in distribution (compared to the flatter distributions of the other collections).

5.8.5 Vault Cranial Sutures

The few significant differences here again centre around Spitalfields, and again are due to the more advanced (older-looking) morphology of Spitalfields compared to Dart and Pretoria. The Grant differences (compared to Pretoria) are more difficult to interpret as actual evidence for differences in ageing rate, as the Grant sample size for vault sutures is low.

5.8.6 Sternal End of the Fourth Rib

While only one instance of significant difference (in distribution between Lisbon and Dart) was found, it should not necessarily be taken as evidence supporting similar ageing rates for the sternal end of the fourth rib, as sample sizes were small. Ribs were often not present or damaged, with no fourth rib sternal ends present.

5.8.7 Distributions of Buckberry-Chamberlain Scored Traits

The results presented provide further, more detailed evidence for population variation in ageing rates. For instance, the transverse organisation and surface texture scores suggest that Spitalfields and Grant are on the "older" end of the ageing scale with regard to these traits. Indeed, where Spitalfields and Grant do show significant differences compared to the Portuguese and South African collections for the other traits, the evidence again points to a relatively advanced ageing rate for Spitalfields and Grant. For surface texture, the fact that very few Dart and Pretoria individuals even reach the highest score of 5 is a further indication of population differences in ageing. That Pretoria and Dart have significantly different score distributions for transverse organisation for the sexes pooled also provides further evidence that such differences can occur despite close geographic proximity. The differences in microporosity distributions give more evidence showing that distributional differences in shape also occur (Dart and Pretoria's relatively flatter distributions compared to those of the other collections).

More important here, though, are the significant differences in Spitalfields compared to the other collections, as Spitalfields was used to develop the age phases and ranges for the Buckberry-Chamberlain method. Spitalfields has significant differences compared to other collections in every scored trait, but particularly and consistently with regard to transverse organisation and surface texture. This implies that there are systematic errors in age estimation that may occur when applying the Buckberry-Chamberlain method to other populations, which have lower ageing rates for some scored traits. Clearly, with archaeological populations, such errors would go undetected, unless it is noticed that particular traits do not appear with great frequency in the target population, despite other age indicators suggesting advanced age. However, as the differences do not consistently occur in every scored trait, some error is perhaps alleviated, as the method relies on combining scores from all traits to give a final estimated age range.

5.9 Within-Collection Distribution Variation by Sex

The results presented for within-collection distribution variation by sex seem to be in line with the results presented earlier for mean ages per phase by sex for each collection – that is, while a few significant results were found, it seems imprudent to suggest that sex-specific standards are necessary based on these few significant differences. Rather, as usual, when assigning age estimates for individuals, all age indicators should be considered to provide the most appropriate final age estimate.

Significant differences in fourth rib phase distribution between males and females of any particular collection are likely due to the paucity in rib data and resulting non-uniformity of age distribution. For example, the Dart Collection female rib data came largely from younger individuals, while the male rib data were skewed towards the older ages.

5.10 Variation in Phase/Score Distribution by Age Group

This part of the analysis was undertaken to examine whether differences in phase distribution (and score distribution for the Buckberry-Chamberlain scored traits) between any collections could be pinpointed as beginning in a particular age group. For instance, as it is known that individual variation in ageing increases with age, due to the increased time for interaction with variables external to the body (culture, lifestyle, diet – factors other than a genetic component) (e.g. see Wittwer-Backofen et al., 2008: 390), significant differences in phase distribution may have started to occur at a specific, presumably middle to older, age group. However, this was not the case; significant differences occur in various age groups, between various collections for various methods, but no discernible trends were found. It is possible that the reason for the lack of an obvious trend was due to the imposition of artificial age groupings – there is no reason to suspect that physiological changes should occur in accordance with strict age-by-decade groupings.

5.11 Mean Age per Phase

Despite some issues with small sample sizes for vault sutures and the fourth rib methods, and unequal variance in some phases for some methods, it is worth noting again that the original sampling strategy was to record equal numbers of individuals of each age category and sex, which generally only became problematic at higher ages (80s and up). This suggests that the inequality in variance and mean between collections in particular phases, and the variation in frequencies of individuals in each phase, is indicative of a real difference between samples in terms of age-related morphological change.

This is true, for example, for the ANOVAs done for Meindl-Lovejoy auricular surface phases. Here, for the 50-60 year phase, the Dart Collection had only 10 individuals, while Spitalfields had 29 with this phase. However, the Dart Collection had the largest total sample size, the oldest individuals (with ages listed over 100 years), and the highest numbers of older individuals – leading to an expectation of higher numbers of individuals with higher scores and at higher phases of the morphological age change, but that is not the case. The low number of individuals scored as having more age-advanced auricular surface scores/phases suggests real differences in rates of ageing compared to other collections sampled. The Pretoria Collection had a similarly low number of individuals in the 50-60 year phase (n = 11). In the 60+ category, the highest phase, Pretoria had only one individual, and Dart had five; interestingly, the Grant Collection had nine individuals presenting the characteristics of this phase, but the lowest total sample size of all the collections. This suggests that Dart and Pretoria have lower rates of auricular surface ageing compared to the other collections.

That some methods show significant differences in sample variance and mean for some phases but not others suggests that samples may differ in the presence of particular morphological traits that were considered diagnostic for an age phase in the reference collection used to develop that method. For example, to belong to a higher phase (and thus, have a higher estimate of age) using the Buckberry-Chamberlain method, high scores in transverse organisation, surface texture, microporosity, macroporosity, and apical change must be awarded. However, if, for a particular population, macroporosity simply is not an age-related feature (or rarely is), leading to individuals having low scores for that trait despite actual advanced age, their age estimates as a whole may be too low as a result of systematic bias due to real, physiological differences in that population's pattern of degeneration – that is, underageing may systematically occur. However, if the trait is age-related, but the morphology resulting in a high score is present at older ages in a target population compared to the reference population, unequal variances and means may occur at the phases in which the trait first becomes highly scored in the reference population, again resulting in systematic underageing of the target population.

Similarly, only phase VI of the Suchey-Brooks pubic symphysis method had a significant difference in means, and this was between Dart and Grant. Figure 5.6 shows that Grant has the lowest mean age for phase VI (at 69.6 years), while Dart has the highest (at 84.3 years). This means that in general, Grant individuals are reaching the advanced morphological stage at earlier ages than are Dart individuals, providing further evidence for Grant's more advanced ageing rate (as discussed in previous sections of this chapter). This also has implications for error when assigning final age estimates to individuals – the phase VI mean for both males and females is around 57 years (although, of course, the range is wide). For example, a Grant Collection individual who was "diagnosed" as phase VI, if given a fairly wide final age estimate of 55 to 70 years at death, would have a chance of being aged correctly, given the Grant mean age of 69.6 years for phase VI (compared to the method's mean of 57 years). However, a phase VI Dart individual given the same final age estimate (based on the mean associated with phase VI as published) would have a much lower chance of being aged correctly, and would likely be nearly 15 years older than the high end of the estimated age range.

When females are tested alone, Lisbon and Spitalfields have significantly different means for phase II. Lisbon's lower mean age indicates that Lisbon females are reaching phase II earlier than Spitalfields females. Lisbon's mean age per phase is lower than that of Spitalfields for phases I to IV, after which, Spitalfields females have the lower mean age (for phases V and VI). This indicates that Lisbon females have a more advanced ageing rate until age-at-death in the 50s, or until phase IV is reached, but the ageing rate slows somewhat in comparison to Spitalfields females, who present the more advanced morphology of phases V and VI at younger ages. These results suggest that not only can population ageing rates differ, but that they may also change over the life course. Indeed, similar "switches" can be seen in Figure 5.6 for the pooled sexes (although differences were not statistically significant); the Dart individuals reach the lower phases at younger ages than Spitalfields individuals for phases I, II, and III, but the ageing rate at older ages seems slower. This is because Spitalfields individuals reach the advanced morphology associated with phases V and VI at younger ages than Dart individuals. Figure 5.7, below, shows the mean ages per Suchey-Brooks phase for females only. Such changes in ageing rate may be the result of differential contributions of underlying factors associated with ageing. As Karasik et al. (2005: 578) note, different activity levels may accompany the varying phases of the life course, subsequently changing the rate of bone activity.

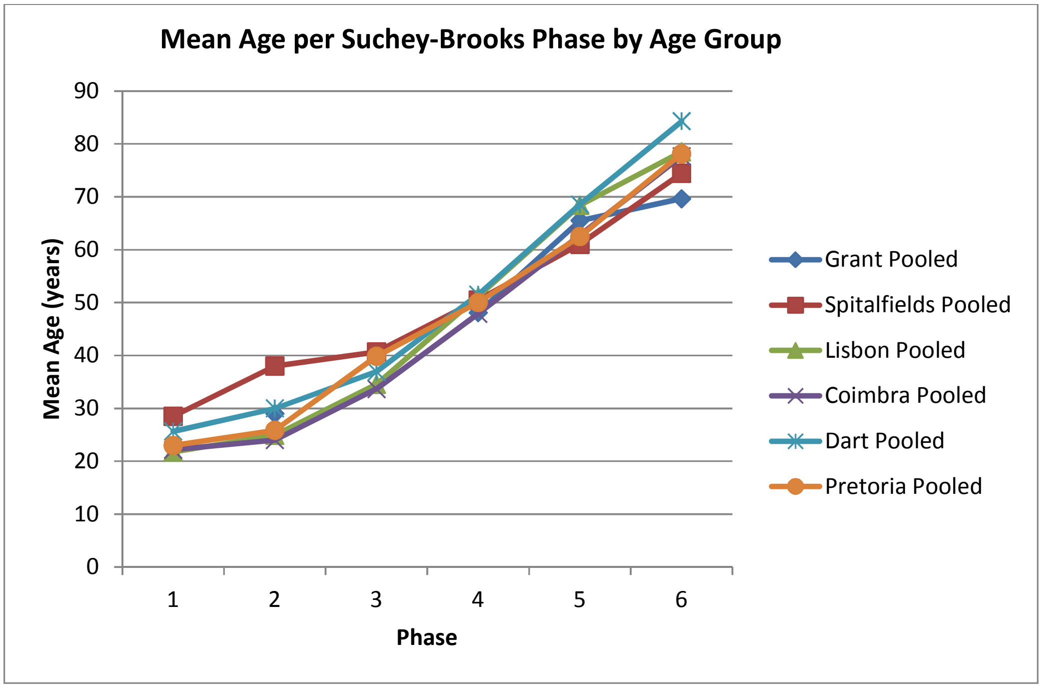


Figure 5.6. Mean age for each Suchey-Brooks phase for all collections, for the sexes pooled

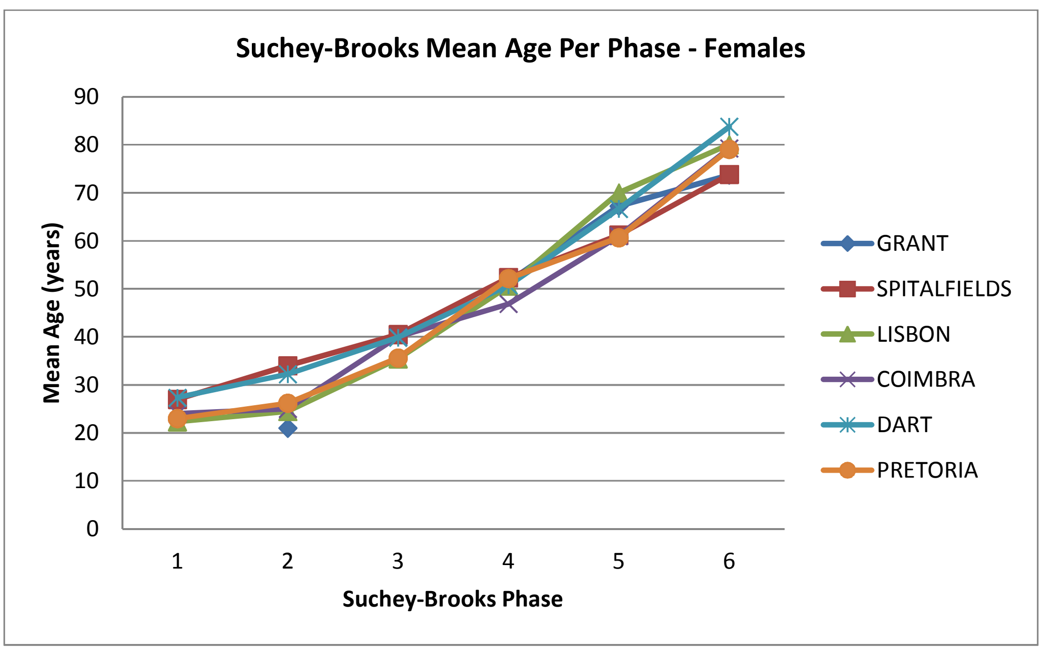


Figure 5.7. Mean age for each Suchey-Brooks phase for all collections, females only

The Meindl-Lovejoy ANOVA results give further evidence in support of population differences in ageing rate – for example, the 40-44 year phase shows significant differences between Lisbon and Spitalfields. Here, Spitalfields mean age is lower than that of Lisbon, suggesting that Lisbon's ageing rate to reach that phase is lower than that of Spitalfields. Similarly, the differences between Dart and Pretoria compared to Spitalfields and Grant suggest that the South African collections have lower rates of ageing than do either Spitalfields or Grant. While not all differences across the phases are significant, there are consistent differences. Pretoria, for example, has higher mean ages per phase for all phases compared to Spitalfields, suggesting a shift in ageing rate so that phases are reached at later ages (a lower ageing rate or slower tempo of ageing). The Buckberry-Chamberlain results again tend to divide the samples by geographic location. The South African collections tend to have higher mean ages per phase, while Spitalfields and Grant tend to have lower mean ages per phase; Coimbra and Lisbon values tend to be intermediate and variable.

The lateral-anterior suture results for mean are somewhat difficult to interpret as so few phases were found to occur in any of the collections. Age ranges for each phase (for all collections) essentially include all of the possible ages sampled, so the mean may not the best measure of difference here. However, mean does mostly increase with phase. Only phase 6 for the sexes pooled shows a significant difference, and post-hoc tests (tests showing where significant differences lie, to be used if ANOVA shows significant difference somewhere) suggest that the differences lie between Grant and Coimbra and Grant and Pretoria. However, the observable Grant lateral-anterior sutures fall in only two phases (6 and 7, except one male in phase 3), and the lower phase has the higher mean age, again casting doubt on reliable interpretations of differences in mean. For males only, a significant difference in phase 6 between Dart and Coimbra again implicates Dart as having the lower ageing rate (higher mean age). Similarly, for vault sutures, age ranges are very wide, although a wider range of phases can be observed. Phase 7 has the most attenuated age ranges (and even these range from the 30s or 40s to the oldest ages), perhaps making it most useful to look for population differences, but no significant differences were found for this phase. Female-only data show a significant difference in mean for phase 6 between Dart and Coimbra (where Dart has the lower ageing rate) and, interestingly, between Dart and Pretoria, where Dart's mean is higher than Pretoria's, indicating that Dart has a slower rate of ageing to reach that phase. While no significant differences were recorded for the fourth rib when sexes were pooled, the significant results for females in phase 10 also must be taken with caution, as the difference is between Lisbon and Pretoria. Only three females in each of these collections were in phase 10. Similarly, significant differences for males in phase 10, suggested to be between Pretoria and Coimbra, and perhaps Spitalfields and

Coimbra, must also be viewed with caution, as there are only five or fewer males in this phase for each of those collections.

In terms of the mean ages per score for the Buckberry-Chamberlain traits, significant differences were also found. For transverse organisation, the only significant difference was in scores of 5 for males only, between Dart and Spitalfields and Pretoria and Spitalfields; as with the results discussed above, the South African means were higher than for Spitalfields, providing further evidence for a lower ageing rate or slower timing in reaching the most advanced phase. The same is true of the significant differences in mean age for surface texture scores of 3 between Dart and Spitalfields. Significant differences in scores of 4 for surface texture of Lisbon compared to both Grant and Spitalfields are due to Lisbon's higher mean, and consequent lower ageing rate. For males only, scores of 3 again show significant differences between the means of Dart and Spitalfields, and, again, Dart's mean is much higher (also because fewer Dart individuals attain higher surface texture scores). Microporosity, macroporosity and apical change also show some significant differences between various collections for all scores, providing more evidence of population differences for particular scored traits; where the South African collections are involved, they tend to have the higher mean ages (and thus lower ageing rates or tempos).

5.12 Trends in Mean Age by Phase

Some methods show significant differences in particular phases between the collections, but not generally in all phases of the method. Figure 5.8 shows the mean ages for each Buckberry-Chamberlain phase – here, only phases 4 and 5 have significantly different mean ages. Each method tested was found to have a significant difference in mean age in at least one phase or category. As noted earlier, these differences in only one phase signify differences in ageing rate. Furthermore, differences that are not significant may still result in small errors in final age estimates.

5.12.1 Auricular Surface

Not many South African individuals presented morphology of the more advanced stages of degeneration in the auricular surface; this was noticed even while collecting the data. More Europeans seem to present advanced-stage morphology, including features such as dense islands of bone or destruction of subchondral bone. These features were simply not common in the South African collections; this is reflected in the low numbers of individuals in the oldest phases of the Buckberry-Chamberlain and Meindl-Lovejoy methods, and particularly in the low numbers of individuals with the highest surface texture scores (for the Buckberry-Chamberlain method). Berg (2008: 570) also noted a difference in the oldest pubic symphyses of females from a forensic Balkan population (victims of recent genocides who were identified through the work of the

International Criminal Tribunal for the Former Yugoslavia, or ICTY) compared to an American sample from the Bass Collection, suggesting that the Balkan females required an additional phase to the Suchey-Brooks system as a result of the “new” oldest-age features.

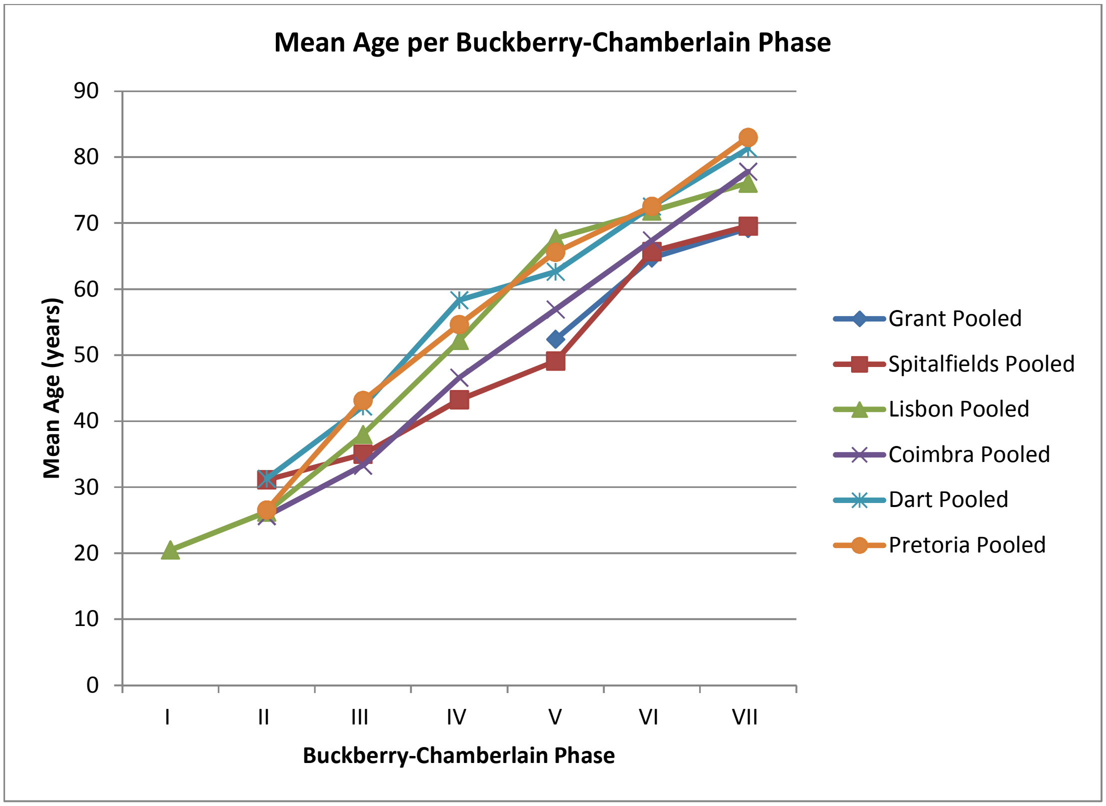


Figure 5.8. Mean age for each Buckberry-Chamberlain phase for all collections, for the sexes pooled

It is difficult to suggest specific reasons for these population differences, given the possibilities: genes, activity, diet, and more. The sacroiliac joint is not particularly mobile; perhaps lower activity levels actually promote degeneration and “older” surface texture morphology. Many South Africans included in the collections were itinerant workers, and presumably many performed manual labour. Conversely, though, work on the pubic symphysis has found a slight link with higher levels of activity resulting in older-looking pubic symphyses (in terms only of ligamentous outgrowth on ventral bevelling), suggesting either higher rates of ageing or faster tempo of ageing (Campanacho et al., 2012: 375). However, the other pubic symphyseal traits examined showed no such activity-ageing rate link, leading the authors to conclude that their evidence could not be taken as support of higher activity resulting in faster ageing rates.

The South African itinerant workers likely had higher activity levels when compared to the Spitalfields individuals, many of whom were involved in the production of lace in some capacity.

Occupations in lace production presumably involved less activity, at least of the lower body. While it is tempting to suggest that these presumed activity level differences are the reason behind the surface texture disparities, it must also be remembered that the named Spitalfields individuals were of the 'middling sort' (Molleson and Cox, 1993). Many of the Spitalfields weavers were managers of the process (master weavers) rather than journeymen weavers who actually worked the looms (Waldron and Cox, 1989: 422), perhaps further highlighting the middle-class status of Spitalfields individuals. The South African itinerant workers would more likely have been of lower status, with the associated possibility of less balanced diets, and cultural differences in diet may also be significant. Thus, diet cannot be discounted as a factor.

As the Spitalfields named sample was largely composed of Huguenots or their descendants, and the Huguenots were a religious minority in France before their flight to England (Cox, 1996: 17-18), it is possible that the group did not mix outside their religion in large proportion, and France and England are far in distance from South Africa (in 18th and 19th century terms). Thus, genes also cannot be discounted as a factor, although their proportional role in ageing is not clear. In any case, what is clear is that there are definite population differences occurring, particularly in terms of surface texture of the auricular surface. This has real implications as to error in final age estimates for populations that do not exhibit the dense islands of bone, or subchondral destruction of the auricular surface that are necessary for placement in the "oldest" phases or scores for auricular surface age estimation methods.

5.13 Overall and Subjective Estimates

It is clear that the "subjective" method of age estimation offers an improvement over simply using the formal methods of age estimation (here, the overall method). However, the Grant Collection exhibited a decrease in the proportion of correct age estimates using the subjective method for the first visit. Interestingly, the proportion of individuals aged correctly with the overall method did not show much change – 38.6% for the first visit, and 35.0% for the second visit. However, the subjective estimates from the second Grant Collection visit did show a marked improvement compared to the overall age estimates and compared to the subjective estimates from the first visit – 50.0% of individuals were aged correctly using the subjective method at the second visit, compared to only 28.9% for the first visit. These changes highlight the impact of observer experience on the accuracy of age estimates. In absolute terms, the number of individuals aged incorrectly using the overall method and correctly using the subjective method were nearly the same for both visits (nine for the first visit and seven for the second visit) – however, these numbers represent 10.8% and 35.0% of the sample, respectively.

When the negative change using the subjective method is taken into account (the individuals aged correctly by the overall method but incorrectly by the subjective method), the second Grant visit again showed a marked improvement. For the first visit, the change was negative – the subjective method was worse than the overall method by 9.6%; for the second visit, the change was positive. The subjective method offered an improvement of 15%. As experience had previously been gained (training and application) in the formal age estimation methods, the results for the overall method did not change much – little intraobserver error was noted at this level (about 3% difference). However, the experience with the subjective method – noticing and interpreting the significance of the informal age indicators and their relationship to age estimated from the formal methods – was gained over the course of the data collection for this project. Here, intraobserver error was much higher, approximately 15%, considering the proportions of correct estimates using the subjective method from the first and second visits, or as high as 25% if the negative changes from overall to subjective methods are taken into account.

These data suggest that user experience is important. Indeed, when the order of data collection is considered, a general trend of increase in the percentage of individuals correctly aged is observed when the negative changes from overall to subjective are taken into account, the exception being the second Grant visit. However, the improvement from first to second visit is of similar magnitude to the changes from overall to subjective in the other collections. This also seems to support the suggestion that experience is important and improves results over time.

The South African collections had the lowest proportions of correct age estimates using the overall method (31.4% for Dart and 39.9% for Pretoria), except for the Grant Collection (38.6% for the first visit and 35.0% for the second visit). The proportions recorded of correct age estimates for the second Grant visit may not be strictly comparable, as the sample size was only 20. The McNemar tests provide statistical evidence of the improvement offered by the subjective method over the overall method – for all collections except the Grant Collection, a statistically significant improvement was observed. Further support for this is in the percentages of improvement (taking into account negative changes). The highest percentages of improvement were for the Dart and Pretoria Collections. Figure 5.9 shows the differences in percentage of correct age estimates for Dart, but it is difficult to interpret the increasing change as being strictly experience-dependent. While Dart and Pretoria were the last two collections sampled, the differences may be because the formal ageing methods worked less well for the Dart and Pretoria Collections; for example, the auricular surface differences in surface texture between the South African collections and the European or European-ancestry collections resulted in auricular surface methods working poorly for older aged South Africans. Probably some combination of experience and variation in ageing and patterns of degeneration are at work.

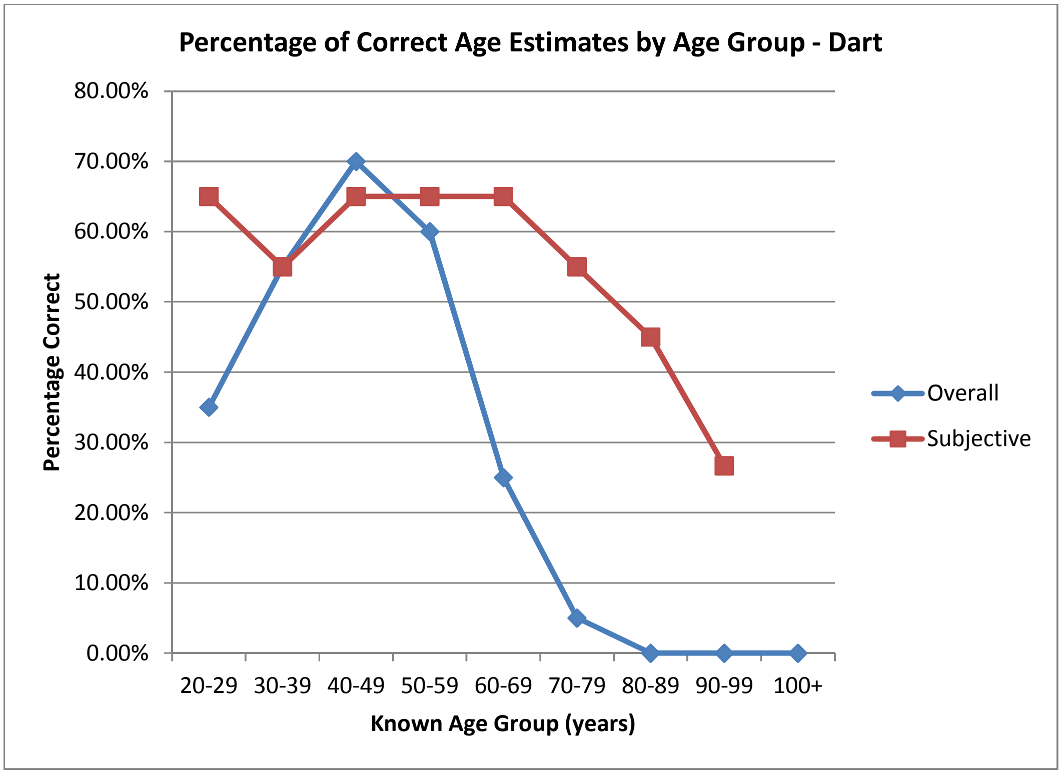


Figure 5.9. Percentages of correct overall and subjective age estimates for each known age group for the Dart Collection

These data provide evidence that subjectivity and reliance on experience need not be viewed negatively. While objective methods would allow for valid statistical comparisons, even supposedly objective methods (metrical sex determination methods, for instance) require some subjective decision making (e.g. the exact placement of skeletal landmarks for measurements). Furthermore, ‘experience and expertise’ are most often relied upon in decisions on age ranges for individuals with skeletal traits spanning more than one phase in a particular age estimation method, and for combining the age ranges of multiple methods to provide a final age estimate (Garvin and Passalacqua, 2012: 430), and were found by Baccino et al. (1999: 932) to improve age estimates. Clearly, experience is already viewed as an asset in osteological work, and is relied upon for formulating an appropriate age estimate; to do so explicitly does not seem an ideological over-extension. Indeed, Baccino et al. (1999: 936) conclude that ‘the most appropriate approach to age estimation should be one that considers all available evidence and recognizes the value of professional training and experience.’

Examination of inaccuracy and bias differences between subjective and overall methods are also informative as to the level of improvement offered by the subjective estimates. The formal age estimation methods (encapsulated in this study by the overall estimates) tend to overage younger adults and underage middle to older aged adults (Bedford et al., 1993: 290; Cox, 2000; Kvaal et al., 1994: 365, 367). The Suchey-Brooks pubic symphysis method tends to begin underageing (as an average) in the 40s (Brooks, 1955: 579; Schmitt, 2004: 2; Martrille et al., 2007: 306; Hens et al., 2008: 1041), although some have reported underageing beginning in the 50s (Sakaue, 2006: 62; Lungmus, 2009: 37), while Saunders et al. (1992) reported underageing from the 30s. Similarly, the Meindl-Lovejoy auricular surface method is generally reported as underageing from the 40s (Lovejoy et al., 1985b: 7; Murray and Murray, 1991: 1167; Schmitt: 2004: 3; Mulhern and Jones, 2005: 63; Martrille et al., 2007: 306; Hens et al., 2008: 1042), although others have reported earlier underageing (late 30s, Osborne et al., 2004: 4) and later underageing (60s, Bedford et al., 1993). Mulhern and Jones (2005: 63) reported that using the Buckberry-Chamberlain method for the Terry and Huntington collections resulted in underageing from the 50s, but for Spitalfields, underageing occurred from the 60s. The fourth rib seems to consistently underage from age 50 and older (Russell et al., 1993: 57; Loth, 1995: 467; Wolff et al., 2012: 374.e7). The cranial suture method begins underageing around 40 years (Brooks, 1955: 580, Figure 6; Wolff et al., 2012: 374.e5), although Lovejoy et al. (1985b: 7) report underageing occurring from age 30 to 39 onwards. While these provide interesting comparisons, the differences may indeed be a consequence of differences between reference and target populations.

In terms of multiple methods, Lovejoy et al. (1985b: 7) note that their summary method results in underageing from ages 30 to 39 and older, while Bedford et al. (1993: 292) found that the summary method began underageing from the 50 to 59 age group. The studies mentioned above generally also found that bias and inaccuracy increase with age (Lovejoy et al., 1985: 7; Bedford et al., 1993: 292; Schmitt, 2004: 2; Mulhern and Jones, 2005: 63; Martrille et al., 2007: 306; Hens et al., 2008: 1041; Lungmus, 2009: 35, 37; Wolff et al., 2012: 374.e5). The results here for the overall method do not differ dramatically from any of the results from other studies: overageing occurs at younger ages, and underageing begins around ages at death in the 50s. This is somewhat later than the single method results above, but in line with the results found by Bedford et al. (1993: 292). As the overall method also includes multiple age estimation methods, this is probably the more appropriate comparison (despite differences in the manner of combination of methods to produce final age estimates). The single methods are still interesting counterpoints; perhaps the use of multiple methods pushes back the age at which underageing occurs.

Underageing also occurs in the subjective estimates, mostly beginning at the 50 to 59 age group, but for Dart and Spitalfields beginning at the 60 to 69 age group. The amount of bias and inaccuracy also increase with age in the results from this study, for both the overall and subjective estimates. However, the amount of both inaccuracy and bias (and standard deviation) are reduced with the subjective estimates compared to the overall estimates, particularly for the oldest age groups, which tend to be the more difficult groups to age. Figure 5.10, below, shows the inaccuracy for Lisbon by age group, while Figure 5.11 presents the bias for Pretoria by age group, both illustrating the improvement offered by the subjective estimates. For all collections except Grant, the mean inaccuracy, bias and standard deviation of the subjective estimates were lower than for the overall estimates. The age groups that were most improved by the subjective estimates in inaccuracy and bias were the youngest two age groups and oldest age groups (from age 70 and over).

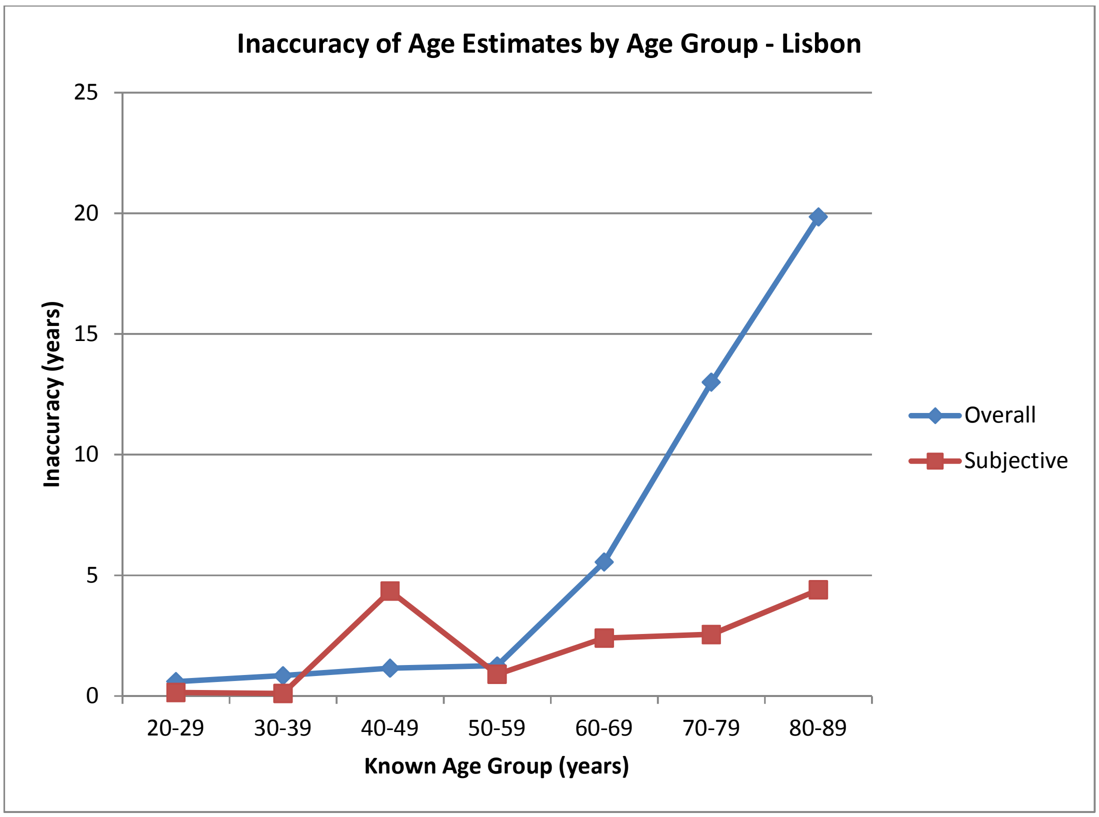


Figure 5.10. Inaccuracy of overall and subjective age estimates by known age group for the Lisbon Collection

While the overall estimate inaccuracy and bias often tended to follow a trend of being somewhat high at the youngest age groups, lowering before the 40s or 50s, and then increasing steadily with age, the subjective estimates did not necessarily (or as obviously) follow such a pattern. The pattern in the overall estimates may be a result of the target population reflecting the age distribution of the reference population (see Bocquet-Appel and Masset, 1982), or a remnant of the statistical procedures involved in developing the method (i.e. regression methods, see Nagar and Hershkovitz, 2004: 151; Aykroyd et al., 1999: 61). While the subjective method still heavily relies on the formal age methods, perhaps it is the addition of the other, informal age indicators and user experience that allows these patterns to be ameliorated.

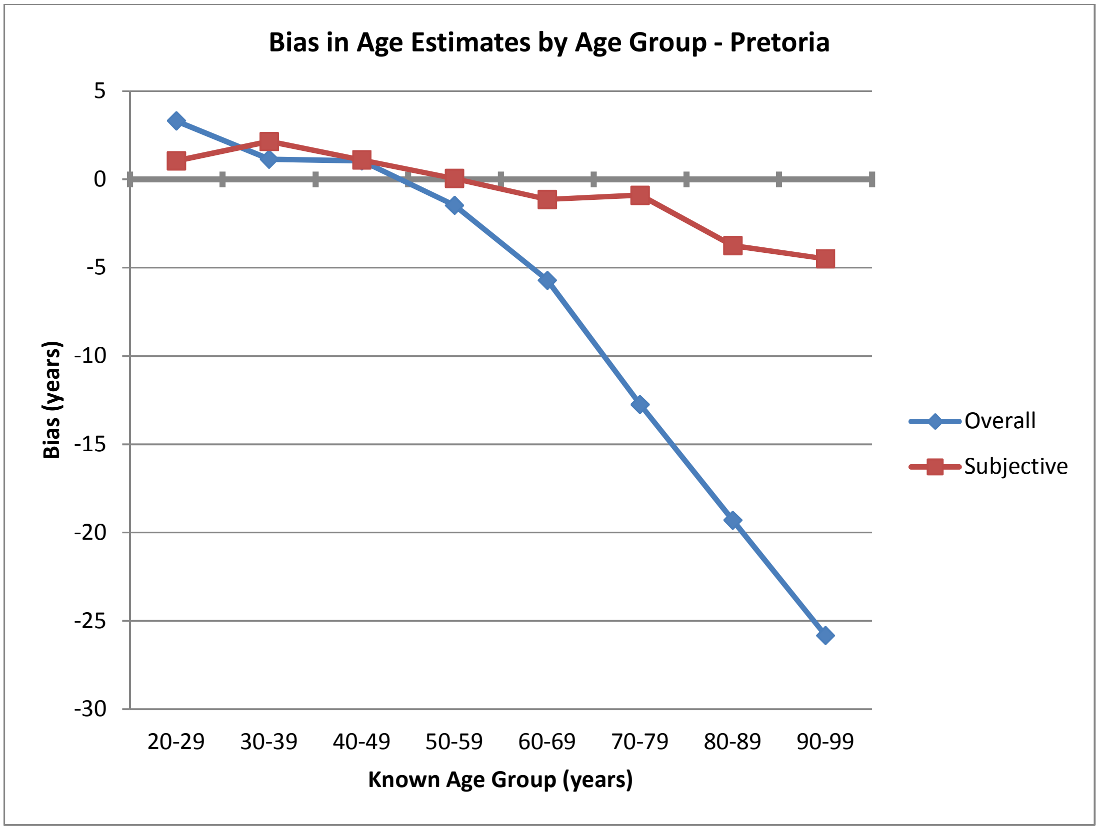


Figure 5.11. Bias of overall and subjective age estimates by known age group for the Pretoria Collection

Evidence of accuracy of age indicators being at least somewhat dependent on experience (Suchey, 1979: 470; Baccino et al., 1999: 935; Saunders et al., 1992; Klepinger et al., 1992: 769), and the possibility of interobserver error (not measurable if the raw data are not available in published studies), alongside the evidence presented here for population variation in ageing rates, makes it seem impractical to compare the levels of inaccuracy and bias found here with that in other studies.

It is well known that the age ranges given as estimates for older individuals are wider than they are for younger individuals (e.g. Wittwer-Backofen, 2008: 384). Certainly this was the case for the overall estimates here. While true to some degree for the subjective estimates, for older individuals these were narrower than were the overall estimates (for the same individuals). The overall estimates for the oldest individuals tended to give a 15-year range, while an effort was made to keep the subjective estimates within a 10-year range, although when age was suspected to be in the 80s or 90s, the age range did tend to widen to 15 years. However, the result was an improvement in accuracy. If older age ranges were widened somewhat for the subjective estimates in accordance with the widths of the overall estimates, further improvement in accuracy may have been seen.

Perhaps more important, though, was the improvement in ability to age the oldest individuals using the subjective estimates over the overall estimates; the oldest ages (generally over age 70) are notoriously difficult to estimate with accuracy (e.g. Berg, 2008: 569). For Spitalfields, no individual over the age of 70 was correctly aged using the overall method, while 21 (out of 37), or 56.8%, were aged correctly using the subjective method. Table 5.1, below, gives the numbers of individuals aged 70 and over correctly aged using the overall and subjective methods. In all cases except for the Grant Collection (which was discussed earlier in terms of correct age assignments and experience/expertise), the subjective method greatly increases the number of individuals over age 70 whose age was correctly identified.

	Total *n* over 70 years	**Overall**		**Subjective**	
		n correct	% correct	*n* correct	% correct
Grant	30	3	10.0%	3	10.0%
Spitalfields	37	0	0%	21	56.8%
Coimbra	40	3	7.5%	18	45.0%
Lisbon	47	0	0%	25	53.2%
Dart	59	1	1.7%	24	40.7%
Pretoria	46	0	0%	27	58.7%

Table 5.1. Numbers of individuals aged 70 and over correctly aged using overall and subjective methods

5.14 Subjective Estimates and Correcting Age Estimates When Ageing Rates Differ

As mentioned above, the Grant, Dart and Pretoria collections had the lowest proportions of correct age estimates using the overall method, consisting solely of the formal age estimation methods. This poor fit of the formal age methods suggests that these collections display differing ageing rates when compared to the other collections. The discussion presented earlier in this chapter suggests that Dart and Pretoria both have slower ageing rates or timing of change than do the European collections, while Grant seems to have a slightly faster ageing rate than the other collections – although the low numbers of young females and 20 to 29 year old males in the Grant Collection may be playing a role here. However, the numbers of Grant males aged 30 and over are comparable to the other collections, and there still seem to be real differences in ageing rate in at least some skeletal indicators.

The subjective age estimates improved on the overall estimates in different ways. The first is that an attempt at narrower age ranges than for overall estimates was somewhat successful. Overall age estimate ranges tended to be 10 to 15 years but wider for the oldest individuals, while subjective age estimates tended to use a range of less than 10 years for the youngest individuals, 10 years for the middle aged individuals and 10 to 15 for the oldest individuals. The second improvement, perhaps more importantly, is that older ages were correctly identified more often using subjective age estimates than overall estimates. Problems in correctly identifying the oldest old when estimating age-at-death of human skeletal remains are widely acknowledged (e.g. Falys and Lewis, 2011). However, here, using all age indicators present, it was more often possible to estimate ages into the 70s, 80s, and 90s, as discussed earlier. By looking for age-related features, such as osteophytes on joints or spicules (new bone formation) in the intertrochanteric fossa of the femur, new bone formation on the margin (and in) the fovea capitis of the femur, sagittal peaking, thinning and angularity of the scapulae, and osteoarthritis and degeneration of joints in general, alongside the formal age estimation methods, it is possible to estimate old ages more accurately.

The fovea capitis changes generally seem to occur after age 40, although in some collections this age was slightly younger or older. As for all age indicators, there is variation, with individuals in their 20s occasionally exhibiting the change, and old individuals occasionally having no changes. The “in-fill” or “overspill” of the fovea capitis seems to occur at older ages – generally over age 55 or 60. Spicules of bone in the intertrochanteric fossa seem to generally occur after age 45 or 50, perhaps becoming more common at older ages (60s and older), although their presence and expression is variable.

The thinning and “scooping” of the parietals, when present, does seem to occur exclusively at the older ages (70s and older). Similarly, thinning of the maxillae, shingle-like ribs and angularity of the lateral border of the scapula also seem to occur exclusively at the older ages (60s and older, but particularly in those aged 70 and older). Overall, however, relatively few instances of these characteristics were recorded so, while these may indeed be good indicators of older age, their absence should certainly not rule out an older age. More systematic research into their rates of occurrence would be beneficial.

The presence of osteophytes on joint margins increases with age, occurring only rarely in the youngest age groups, and tending to become more common (and on multiple joints) in the 50 to 59 age group, although this is variable. For Dart and Pretoria, for instance, the late 30s and early 40s were the ages for common osteophyte occurrence. Vertebral osteophytes may be a particularly good informal indicator of age, becoming more common in the late 40s and early 50s. The vertebrae were the most common location for osteophytes across the collections sampled here, although their absence alone should not be considered evidence for placement in a younger age group. The presence of multiple joints showing degeneration (osteophytes alone and OA) seems to be a good indicator of older age (when trauma cannot be implicated as a cause);

particularly, multiple joints with OA (and osteophytes alone) seem to occur in individuals aged 80 and over.

Most of these informal age-related features have not been extensively researched, although an early reference to sagittal peaking (thinning and scooping of the parietals) as being associated with old age was reported (Ferré, 1876: 423-424). However, some research on the relationship between degeneration of the various joints with age can be found. The appearance of osteophytes at joint margins and new bone formation at muscle attachment sites, the former being described by Scheuer (2002: 306) as the beginning of degenerative joint disease, tend not to occur until after 40 years of age. This seems to agree with the osteophyte data presented here. However, this is variable because of contributory factors such as nutrition, lifestyle and genes (Scheuer, 2002: 306).

Previous studies have found that the presence of vertebral osteophytes is age-related, although individual variation and the inability to attribute osteophytes strictly to age because of underlying causative factors (including mechanical wear and tear and an inherited predisposition) does not make it appropriate for development into a formal age estimation technique (Watanabe and Terazawa, 2006:159; Listi and Manhein, 2012: 1539). Instead, it might be used for general age estimates. Nathan (1962: 245-246) found that small osteophytes occurred in every individual by the early 40s in the documented Hamann-Todd collection sample, and the size of the osteophytes increased with age. Listi and Manhein (2012: 1539) also found a correlation between vertebral (facet) OA and age, and between vertebral osteophytes (on the bodies and facets) and age.

Knee and hip osteophytes have been found to occur slightly earlier in males compared to females, and the presence of osteophytes does not necessitate subsequent development of OA; rather, osteophytes are suggested to be related to "normal" ageing processes (Danielsson, 1964: 13; Danielsson and Hernborg, 1970: 312; Hernborg and Nilsson, 1973: 73; Karasik et al., 2005: 578). However, other clinical research suggests that while osteophytes are a common occurrence in individuals who develop OA later, in these individuals, osteophytes increase in size at a faster rate than in individuals who do not develop OA (Hernborg and Nilsson, 1973: 73).

Risk factors for OA include sex, "race", age, genes, occupation (activity), and previous trauma, but age 'is arguably the most important', while the other risk factors will affect the rate of acceleration of age-related degeneration (Kirkwood, 1997: 683; Toivanen et al., 2010; Anderson and Loeser, 2010; Salter and Lee, 2012). The exact mechanisms of OA are not clearly understood. It remains to be seen as to whether OA represents a deviation from the path of "normal" age-related cartilaginous degeneration, or whether it is simply the end point of “normal” degeneration (Kirkwood, 1997: 691). While OA may occur at younger ages, it seems to be nearly universal in individuals in their late 60s and older. It obviously does not occur in the same joint in every individual at least partly due to differences in activity (level and type), but it does seem to occur to some degree (either porosity plus marginal osteophytes, or eburnation) in the majority of individuals at older ages. While it is possible for a young individual to exhibit osteoarthritis, a traumatic origin is possible, which might be evidenced in the skeleton. Furthermore, the formal age estimation methods would hopefully indicate the younger age of the individual. Forensic anthropologists also use the presence of osteoarthritis in joints and vertebrae and ‘bone quality’ as age indicators for final age estimates (Garvin and Passalacqua, 2012: 429).

It seems that the informal age indicators discussed in this section hold some promise as additional age information, particularly when previous studies have provided support of their relationship with age (as in the short discussion of osteophytes, above). Indeed, that the subjective estimates, taking these indicators into account, performed better than the overall estimates, which relied solely on the formal age estimation methods, is good starting evidence for their application on other populations. As usual, caution is required, as absence should not be taken as positive evidence for admission into some age category or another, but the results here do support their use as additional age information that can aid in providing more accurate age estimations (when they are present).

If it is observed that a target sample does not seem to exhibit high scores for a particular feature (for example, as discussed earlier, the Dart and Pretoria Collections had very few high surface texture scores for the auricular surface), an estimated age distribution could first be constructed, and mean ages per phase or score calculated based on the estimates. These could then be compared to published mean ages per phase of the reference sample. If it is clear that there is a particular phase where the difference between target and reference sample mean ages widens, then the absolute difference can be calculated and then added to each individual age in the target sample. Alternatively, a percentage difference can be calculated from the published means or perhaps the high end of the published age ranges, and that percentage can be added to an individual’s final age estimates where appropriate. Careful reading of the published written descriptions accompanying the instructions for scoring phases or scores in the methods used and comparing them to the feature’s presentation in the target sample will also help in deciding at which phase corrections should be applied.

5.15 Variability in Ageing Rates: Summary

Hoppa (2000: 186) found differences in pubic symphyseal ageing rates particularly in females over 30 years of age between a reference sample (the 20th C County of Los Angeles forensic sample used to develop the Suchey-Brooks pubic symphysis method; see Katz and Suchey, 1986) and two target populations (Spitalfields and another County of Los Angeles forensic sample drawn from the same collection as Katz and Suchey, 1986; see Molleson and Cox, 1993, and Klepinger et al., 1992). However, for this study, Hoppa (2000) did not examine the bones himself; rather, the raw data was given to Hoppa by Suchey, Klepinger and Molleson. Interobserver error was not taken into account. Most pubic symphyses do not conform exactly to the photographs published with instructions for the methods or the casts, which are meant to represent each phase. Rather, individual pubic symphyses might have bony features from more than one phase, or appear intermediate in morphology. The researcher must decide which phase fits best with the morphology presented; different researchers may emphasise different bony features in making phase decisions, and therefore make different decisions.

This inherent subjectivity (interobserver error) in the application of age and sex methods makes comparison between studies inappropriate when the research goal is to examine variability in ageing or sexual dimorphism. Of course, there are often political issues with the curation of human remains, necessitating repatriation and/or reburial and impeding further analysis by different researchers; at other times, accessibility of the collections may be dependent on the research plan, or costs of research may be prohibitive. Thus, comparison of data collected by different researchers can be necessary, as bioarchaeologists must work with what is available. However, for research questions regarding efficacy of methods and ageing rate and sexual dimorphism variation, such comparisons should not be attempted.

Some (Nagar and Hershkovitz, 2004: 151; Aykroyd et al., 1999: 61) have noted that the statistics used to develop methods may result in specific biases. Specifically, logistic regression may be at least partially to blame for the typical bias pattern of overageing of younger individuals and underageing of older individuals. In logistic regression, a line is fitted to data, leaving the middle portion of the line more closely approximating the mean than the ends (the best fit is determined by the age distribution of the reference population). If this is true, it may be possible to detect such an effect in the distributions of correct ageing – overall and single-method age estimates should be correct more often for the middle ages than younger and older ages. If this pattern does not occur, differences in proportions of correct estimates are more likely due to other causes, including real variation in ageing rates. However, most researchers (e.g. Wittwer-Backofen et al., 2008: 390) note that wider variation in ageing occurs in older individuals, who have had more life-years to accumulate the differential effects of activity, diet and any other factors on bone. As such, if younger and middle ages are more often correct than older ages, either differences in ageing rates or methodological issues (i.e. logistic regression used in method development) could be the reason.

Another source of bias may be the reference sample used to develop an ageing method, as age distributions for the target samples on which the method is subsequently applied may mimic the age distribution of the reference sample, reflecting the method's construction (Bocquet-Appel and Masset, 1982). In this study, other reference sample problems have been observed in that the reference sample may exhibit a somewhat unique suite of skeletal characteristics used in the scoring of the method, or a different ageing rate to that of the target populations, resulting in error in final age estimates. Here, the Spitalfields sample was found to display surface texture characteristics in the auricular surface (in older individuals) and score distributions significantly different from that of most other collections sampled (particularly when compared to the South African collections). As Spitalfields was used in the development of the Buckberry-Chamberlain auricular surface method, when applied to target samples lacking this older-age surface texture morphology, errors in age estimate may result. Similarly, although the Meindl-Lovejoy auricular surface method was developed on a mixed sample, including individuals from the Hamann-Todd Collection, individuals from the archaeological Libben Collection, and modern forensic cases, it seems that these Americans also displayed the advanced morphological surface texture characteristics that were not often observed in the South African collections. Thus, as few of the South Africans from this study were placed in the oldest Meindl-Lovejoy phase, for which surface texture (subchondral destruction) is a key characteristic, the result is again error in final age estimates – few older individuals were aged correctly using the formal age estimation methods.

Meindl and Lovejoy's ectocranial suture closure age estimation method, also tested here, was developed on the Hamann-Todd Collection. For this method, it was more difficult to interpret possible differences in light of the reference collection, as the sutures were deemed best for estimating age in general terms, and not suitable for specific, narrow age ranges. Similarly, sample sizes available to this study for the fourth rib age estimation method, developed on an autopsy sample from the Broward County Medical Examiner's Office, Florida, USA (İşcan et al., 1985: 853), were generally too small to generate any interpretation on population differences and the effect of the reference sample.

Meanwhile, relatively more significant differences between collections in phase or score for any particular age group may suggest locations across the life course for any differences in ageing rates, or ages at which individual variation is highest.

5.15.1 Transition and Bayesian Analysis: A Solution?

Transition analysis and Bayesian methods have been offered as solutions to the age-mimicry problems in the application of age standards developed on reference populations to target populations with different age distributions. Bayesian methods require the use of a prior age distribution, chosen by the researcher – either a specific informative prior (with an attempt at matching the age distribution of the target sample), or a uniform prior. However, when using transition analysis with Bayesian methods, informative priors with ageing rates similar to that of the target population provide better results than do informative priors with ageing rates dissimilar to the target population (Godde and Hens, 2012: 263). Godde and Hens (2012: 264) also note that while 'the results from Bayesian analysis indicate that careful selection needs to be maintained on the within phase age-at-death distribution in transition analysis', '[C]onversely, the selection of a uniform prior can be relaxed as long as transition analysis sample age-at-death distribution matches the target sample.' Of course, the problem with target samples is that the age-at-death distribution is not usually known. The within-phase age-at-death distribution they mention was analysed by Kolmogorov-Smirnov tests of the age distribution (presumably in decade-long sections) for each phase; here, ANOVAs were used instead, but the results both provide evidence for differences in ageing rates between populations. Parametric tests (such as ANOVA) tend to be preferred over nonparametric tests (such as the K-S test) due to robusticity and strength. Godde and Hens (2012: 260, 263-264) found differences in ageing rates between their Sardinian historical target sample (from the Sassari Collection, curated at the University of Bologna, Italy) and a Balkan genocide reference sample (again documented by the work of the ICTY). Interestingly, fewer significant differences were found between the Sardinian sample and a Terry Collection sample, despite wider geographical distance between Italy and the USA compared to Italy and Kosovo and Croatia. While transition analysis and Bayesian methods may result in higher accuracy than the use of single ageing techniques alone (as was found with test of the Suchey-Brooks method by Godde and Hens), knowing whether a sample used for transition analysis has a similar ageing rate to that of the target sample is problematic, and with no immediate solution (Godde and Hens, 2012: 264). The use of model prior probabilities offers one potential solution, and using a number of different model prior probabilities to evaluate the impact on the various outcomes seems currently to be one potentially useful approach (Gowland and Chamberlain, 2005).

As found in other studies, some ageing methods simply do not perform very well. For example, cranial suture closure shows only a very small increase in mean age by phase. Accompanying this are very wide age ranges, which was also found for the fourth rib method; fourth rib changes also do not consistently increase in mean age by phase – however, this could partly be due to low numbers of individuals with preserved fourth ribs. This suggests that if bioarchaeologists and forensic anthropologists are going to use cranial sutures to estimate age, it is probably best to use them only to suggest broad life stages, such as young adult, middle adult and old adult.

For the pubic symphysis and both auricular surface methods, the mean ages per phase generally increase consistently and range less widely than for cranial sutures and the fourth rib.

Because every attempt was made to have a uniform age distribution, with 10 individuals per sex per known age group where possible, the age distribution should not affect the distribution of scores/phases. Thus, where there are differences in mean age per phase, it reflects variation in ageing rates.

The results do show differences in mean age per phase or score by collection. When collections from the same country are compared (for example, Lisbon and Coimbra or Dart and Pretoria), the within-country mean ages are more similar than mean age per phase between countries. This holds for age indicators that do indeed show morphological change with age (but less so for cranial suture closure). Furthermore, some differences are statistically significant, but not necessarily across all phases used by a method. For example, phase VI of the Suchey-Brooks method presented statistically significant differences in mean age, but the other phases did not.

5.15.2 Phase/Score Differences

The differences in distributions of phases or scores were also informative; such differences are further evidence of variation in ageing between populations. There were statistically significant differences between the Suchey-Brooks phase distributions between Grant and Lisbon and Spitalfields and Grant. For the Meindl-Lovejoy auricular surface method, the Grant Collection and Spitalfields were different from every other collection, including each other. Similarly, for the Buckberry-Chamberlain auricular surface method, the Grant Collection's phase distribution was significantly different from every other collection. For the Buckberry-Chamberlain auricular surface method and lateral-anterior cranial sutures, Spitalfields showed significant differences in phase distribution with every other collection except Coimbra. Only a few significant differences were found with vault sutures – between Spitalfields compared to both Dart and Pretoria, and between Grant and Pretoria. For the fourth rib, the only significant difference in distribution of scores was between Dart and Lisbon; however, sample sizes for the fourth rib were low for all collections.

Because the Buckberry-Chamberlain method requires scoring of each feature to get a composite score, it was possible to analyse score distributions for each feature. While each feature did have some significant differences in distribution between various collections, surface texture distribution was significantly different for Spitalfields compared to all other collections except Grant. This is particularly important because Spitalfields was the reference collection used to develop the Buckberry-Chamberlain method. That the surface texture score distribution for this collection was significantly different from nearly all the other collections tested has implications for the method's application to other skeletal collections – error is practically in-built in the method, as most other collections did not have such high frequencies of "advanced" auricular surface ageing in terms of surface texture, particularly the South African collections.

In general, inaccuracy of both the overall and subjective estimates increases with age, which is no surprise, as other studies have found the same. This is because individual variation increases with age, as mentioned earlier.

5.15.3 User Experience

In terms of the overall estimates, much like in other studies (Bedford et al., 1993: 290; Cox, 2000; Kvaal et al., 1994: 365, 367), overageing occurs at younger ages, and underageing begins around ages at death in the 50s and 60s. Compared to the overall estimates, the subjective estimates improved accuracy in all but the Grant Collection – and that was the first collection visited. The Grant Collection was revisited to perform an intraobserver error test, and the subjective estimates improved. Generally, accuracy improved by about 20% compared to the overall estimates when using subjective estimates. Furthermore, the mean bias was much less with the subjective estimates than the overall estimates, and neither bias nor standard deviation showed the dramatic increase with age as with the overall estimates.

The success of the subjective age estimates are user-experience dependent; this can be illustrated in the intraobserver error sample study of the Grant Collection. The second sample consisted of 20 individuals, while the original sample was 83. There was about a year between the first and second sampling. The accuracy of the overall estimates, based only on the formal methods, remained basically the same, with only a 3.6% difference. However, the accuracy of the subjective estimates improved dramatically, from 29% to 50%. In the year between samplings, experience was gained by collecting data from over 700 other skeletons. This experience improved the application of subjective age estimation.

This highlights the need for experience in this type of research. Researchers need more knowledge of the normal range of human variation, ageing, and sexual dimorphism. While statistical alternatives to morphological age estimation have been presented to try to circumvent variation, they necessitate assumptions, some of which the results presented here negate. For example, ageing rates vary but the shape of the distribution of scores is the same across populations. However, even the assumption that a phase is conditional on age should be entertained with caution, particularly for cranial suture closure. In more statistical terms, significant differences in central tendency and shape of distributions have been found. This provides evidence that variation in ageing rates and sexual dimorphism can be reflected in the rate or timing of change, or in the expression or frequency of age-related characteristics.

This means that if a reference sample used to develop a method has a unique rate of change for a feature scored in a method, any application of the method to other populations will result in error in age estimates. Of course, this makes things difficult and age estimates less accurate. However, it suggests that subjectivity in deciding what features to weigh more heavily is appropriate in some cases. For instance, when scoring the South African auricular surfaces, it was noticed that the dense bone that is necessary for high Buckberry-Chamberlain surface texture scores did not often occur, regardless of other indicators suggesting advanced age. As Figure 5.12 shows, surface texture scores of 5 were not common amongst the Dart or Pretoria individuals. Similarly, for Meindl and Lovejoy's oldest age group, 60+, there are not many South Africans, likely largely as a result of a lack of subchondral bone destruction – in other words, surface texture. Such observations can be used to adjust age estimates accordingly, and obtain more accurate age estimates.

Other skeletal elements, such as the pubic symphysis, seem to present fairly similar changes across populations – at least of the collections studied here. This suggests that it is appropriate to give more weight to pubic symphysis age data, for instance, which is what bioarchaeologists tend to do in estimating age-at-death. As is often suggested, using as many methods as possible should also help (Todd, 1920: 314; Todd and Lyon, 1924: 380; Acsádi and Nemeskéri, 1970: 120; Meindl and Lovejoy, 1985: 65-66; Brooks and Suchey, 1990: 237; Buikstra and Konigsberg, 1985: 318-319).

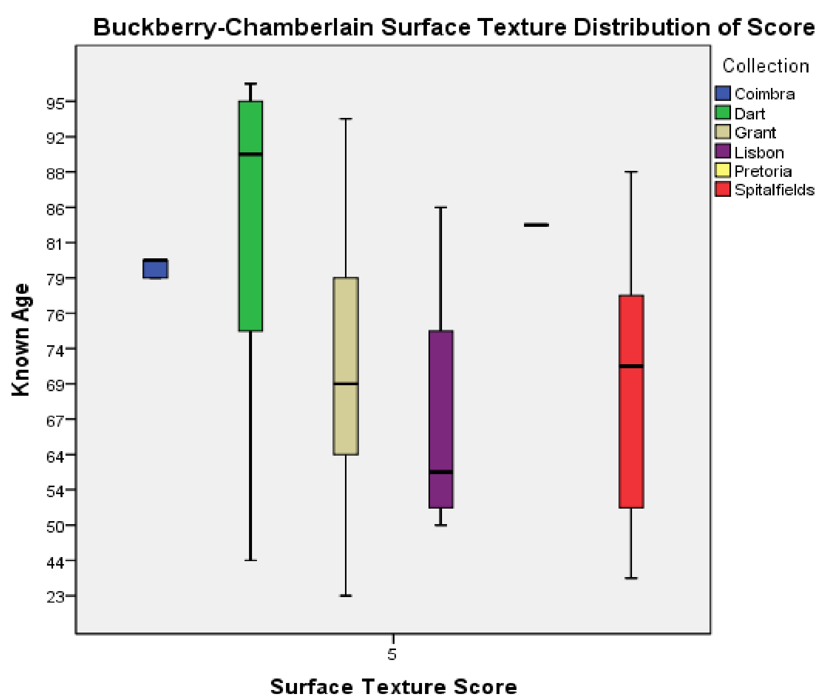


Figure 5.12. Boxplot of Buckberry-Chamberlain surface texture distribution for scores of 5

Such subjectivity is similarly sometimes necessary in assessing sexual dimorphism; if a particular target population has a certain feature with an obviously different scale of sexual dimorphism from the scale presented in the sex determination method, adjustments must be made. For example, males and females from a Scottish population from Orkney were found to have very wide, square mental eminences, and it was noted that adjustments had to be made in order to make appropriate sex determinations when isolated mandibles were found (Gooney, 2012). This is because other skeletal evidence had clearly shown that the individuals with wide, square mental eminences were not all male (Gooney, 2012). Clearly, here, experience was important, and the adjustments made were subjective.

By using all available age and sex information, including subjective indicators, it is possible to account for population variation to arrive at more reliable age and sex determinations, and therefore, stronger bioarchaeological and forensic interpretations.

5.16 Inter- and Intra-observer Error

5.16.1 Sex

The intraobserver error found by Rogers and Saunders (1994: 1051), in their study of morphological traits of the pelvis used for sex determination ranged from 0 to 11.3%, although a suite of 17 traits were examined. Relevant to this study (with associated intraobserver error rates) were the following: ventral arc: 0%; subpubic concavity: 3.2%; sciatic notch: 6.5%; ischiopubic ramus ridge: 11.3%. As with the Williams and Rogers (2006) study, the sex data were only recorded with three options (male, female and indeterminate) (Rogers and Saunders, 1994: 1048), so the intraobserver error rates record disagreement. Here, the most reliably scored trait was the ventral arc, followed in order by the subpubic concavity, ischiopubic ramus ridge and greater sciatic notch. This is nearly the same as that found by Rogers and Saunders (1994), with the exception here that the ischiopubic ramus ridge was more reliably scored than the sciatic notch, but vice versa for Rogers and Saunders. Intraobserver error rates were much higher here, for possible reasons that will be discussed below.

Williams and Rogers (2006: 730), in their study of morphological cranial traits used for sex determination, found intraobserver error of 8% for the sexes pooled in terms of sex determination. The intraobserver error rates for the individual skeletal traits ranged from 2% to 20% (they excluded any trait with intraobserver error of 10% or greater, as this was the set critical value). A sample of 50 individuals (half male, half female) from the Bass Collection was used (Williams and Rogers, 2006: 730). The traits similar to those observed here had intraobserver error rates as follows: chin form: 2.0%; size of supraorbital ridge: 6.0%; size of mastoid process:

8.0%; orbital margins: 10.0% (Williams and Rogers, 2006: 732). The sex data were only recorded with three options (male, female and indeterminate), so intraobserver error rates are only in disagreement rather than distance in scores from first and second visit, as was done here. Rogers (2005: 497) reported intraobserver error of 12.2%, and accordingly suggested that some traits were difficult to consistently score. The individual traits also used in this study, with their associated intraobserver error are as follows: mastoid process: 0.0%; supraorbital ridge (glabella): 2.0%; nuchal crest: 2.0%; chin form (mental eminence): 4.1%. In this study, the lowest rates of intraobserver error were for the mental eminence and glabella, followed in order by the nuchal crest, supraorbital margin and mastoid process, suggesting that the mastoid process is most difficult to score. Here, then, the traits that were more reliably scored were different to those found by Rogers (2005), although Williams and Rogers (2006) also found that mental eminence was reliably scored.

Overall, intraobserver error rates in this study were higher; experience in following the written descriptions likely played a large role here. However, had only three possible scores been used here, it is possible that intraobserver error scores would be lower, as it seems likely that the qualitative aspects of Williams and Rogers' scores would make it less likely that a score change would occur. Here, though, where scores of 1 and 2 are both essentially female, if category change occurred between intraobserver error visits (score of 1 on the first visit, but score of 2 on the second visit), a disagreement nevertheless occurred. As a comparison, when intraobserver error for the overall sex determinations was examined here, only in one case was there a change from "female" to "male" between visits; all other differences were in degree of sexualisation – that is, for example, from "probable female" to "female". This lends some support to the idea that had fewer score categories been used for each skeletal trait, as in Williams and Rogers, intraobserver error would likely have been lower.

Intraobserver error results for Albanese's method were good: no change in sex estimation occurred between visits, and the error rates for the measurements themselves were uniformly low, suggesting that the landmarks outlined by Albanese (2003a) are reliable and easy to reproduce.

5.16.2 Age

The intraobserver error results for the age estimation method were somewhat disappointing; the Suchey-Brooks method had the best results, and even here, there was phase agreement in only 54% of individuals. However, for both Suchey-Brooks and Meindl-Lovejoy, the phase difference was less than one phase on average. The phase difference between visits was slightly higher for the Buckberry-Chamberlain method, at 1.13. In terms of the final age estimates, the overall estimates did not change much between visits, while the subjective age estimates were noticeably improved on the second visit. This further underlines the fact that experience is important in the use of the informal age indicators. It is interesting, though, that despite the fairly high error rate for each formal method independently, the overall final age estimates, which relied solely on these formal methods, did not change much on average. This suggests that perhaps the phase difference between visits (less than one for Suchey-Brooks and Meindl-Lovejoy, and just over one for Buckberry-Chamberlain) may be more important in determining the reliability of a method than simply observing the phase agreement across visits.

Interobserver error results suggest that the Suchey-Brooks method is easier to replicate than the Meindl-Lovejoy auricular surface method. More cases of agreement occurred with the Suchey-Brooks method (although it was still low, at or near 50% for Coimbra and Spitalfields), but phase differences between observers was also lower compared to Meindl-Lovejoy. While this may be a function of fewer possible phases in the Suchey-Brooks method, it would be hoped that any disagreement over phase would be within one phase up or down of the initial (first observer) phase placement. However, the higher Meindl-Lovejoy phase differences (all over one phase) suggest that this was not always the case, while all Suchey-Brooks phase differences were lower than one (some phase differences were only a half phase difference, where one observer placed an individual in a range of phases – e.g. 1 to 2).

For the Buckberry-Chamberlain scores for each trait, reliability seems better than for the Meindl-Lovejoy method, as for each trait, agreement was over 50%, and phase differences are all lower than one. Again, this is the result of some differences in the scoring of an individual's trait by less than one, where one observer has scored the individual across a range of phases, or as a half score (e.g. a microporosity score of 1 to 2, where it is perhaps not entirely clear where the "boundary" for the microporosity lies in terms of percentage of surface covered). Some disagreement may also result from differences in descriptions of traits; Dr. Gowland tested the Buckberry-Chamberlain method before it was published, when the scoring system began at 0 instead of 1, while the current author used the published version, where the scoring system begins at 1. While it is not known if there were differences in the written descriptions, it is a possibility that some changes or improvements were made prior to publication. In general, the increased reliability of the Buckberry-Chamberlain method over the Meindl-Lovejoy method suggests that the former is easier to use.

Unfortunately, it is difficult to interpret whether experience also plays a role in interobserver error in this case because, while Dr. Gowland certainly has more experience and expertise than the current author, the data were collected as part of Dr. Gowland's post-doctoral research – so the cautious researcher would have to surmise relatively equal experience between observers.

5.17 Key Findings

In this section, the most important points from the discussion of the results are highlighted. These include the most or least sexually dimorphic single traits used in sex determination, the relative reliability of the age estimation methods, variation in ageing rates and expression of sex determination, and the advantages the subjective method of age estimation provides compared to the overall method. These key findings directly address the research aims stated in Chapter 1: to quantify variation in skeletal ageing rates and expression of sexual dimorphism and to assess the relationship of skeletal indicators over the life course.

5.17.1 Sex Determination

5.17.1.1 Sex Determination Using Specific Features of the Skull

- the least sexually dimorphic single indicators were the mastoid process and supraorbital margin
- the mental eminence was also not found to be a highly sexually dimorphic trait
- the glabella was not as sexually dimorphic for the South African collections compared to the other collections, but was the most sexually dimorphic single indicator for Spitalfields and Coimbra

5.17.1.2 Sex Determination Using Specific Features of the Pelvis

- the sciatic notch was the least sexually dimorphic of the single pelvic indicators
- the most sexually dimorphic (and thus potentially most useful in work with fragmentary remains) single pelvic indicators were the subpubic concavity and the ventral arc, followed by the ischiopubic ramus ridge, with clearly bimodal score distributions
- scores for pelvic traits are generally far less variable than for any of the single skull traits used to determine sex

5.17.1.3 Albanese's Metrical Method

- modification 1 is best for determining sex compared to the other modifications available
- within-country similarities tended to be present, supporting the extrapolation of methods developed using a particular reference collection to skeletons of similar geographic origin

5. 17.2 Variation in Expression of Sexual Dimorphism

- females were more often sexed correctly than males when determining sex using the skull, suggesting that all of the collections analysed (with the exception of Grant) exhibit a scale of sexual dimorphism of the skull that is skewed towards the "female" end of the spectrum compared to the skeletons used to develop the skull scoring method
- within-country results for correct sex determinations using the skull were more similar than results between other collections
- this evidence makes it clear that while some traits may cluster in terms of expression of sexual dimorphism between populations that are close geographically (e.g. mastoid process clustering of results between Dart/Pretoria and Coimbra/Lisbon and glabella clustering in Dart and Pretoria), other traits may exhibit significantly different ranges of expression between the same two populations (e.g. significant differences in the expression of sexual dimorphism in the supraorbital margin and the nuchal crest between Coimbra and Lisbon)
- with regard to sexual dimorphism in the femoral and pelvic measurements used in Albanese's metrical method, within-country similarities are found, but not every measurement shows such within-country similarity in the scale of sexual dimorphism (e.g. SPRL clusters for Pretoria/Dart, but not Lisbon/Coimbra)
- clustering of percentages of correct sex determination using the skull provides further evidence for regional and perhaps temporal variation in sexual dimorphism
- the results presented here do not support either Meindl et al.'s (1985b: 84) or Mays and Cox's (2002: 125) conclusions (respectively, that males are more often misclassified than females, or vice versa)
- the suggestion that using multiple indicators is best for sex determination (e.g. Meindl et al., 1985b: 85) is supported
- in terms of sex determination, no age-related trends were found in any of the collections under study here (e.g. older females and younger males seemed no more likely to be misclassified than younger females and older males: Walker, 1995, 2005)
- evidence found here highlights the variability in sexual dimorphism across populations

5.17.3 Age Estimation

- Suchey-Brooks method is easier to replicate than the Meindl-Lovejoy auricular surface method
- it is appropriate to give more weight to pubic symphysis age data in age-at-death estimation

- reliability was better for the scoring of the Buckberry-Chamberlain traits compared to the Meindl-Lovejoy method
- because no consistent differences or trends between males and females were found in the Buckberry-Chamberlain scored traits, the evidence seems insufficient to suggest separate male and female scoring standards

5.17.4 Variation in Ageing Rates

- the evidence suggests that Dart and Pretoria have slower ageing rates or timing of change than the European collections, while Grant seems to have a slightly faster ageing rate than the other collections
- the differences in skeletal characteristics and qualities emphasised in each of the auricular surface methods (i.e. Meindl-Lovejoy and Buckberry-Chamberlain) should be noted; these differences may be the result of different morphological expression of ageing in the auricular surface of the different reference collections used in each method's development
- for any particular ageing method, collections or populations may differ in the expression or presence of particular morphological traits that were considered diagnostic for an age phase in the reference collection used to develop that method
    - e.g. not many South African individuals presented morphology of the more advanced stages of degeneration in the auricular surface despite older age
- in terms of the Buckberry-Chamberlain auricular surface method, the significant differences in Spitalfields (i.e. in surface texture) compared to the other collections have important implications, as Spitalfields was used to develop the age phases and ranges for the Buckberry-Chamberlain method
    - systematic errors in age estimation may occur when applying the Buckberry-Chamberlain method to other populations
- population ageing rates differ, but that they may also change over the life course

5.17.5 Overall Method vs Subjective Method

- user experience is important in the use of the formal age estimation methods and the informal age indicators
- even supposedly objective methods (metrical sex determination methods, for instance) require some subjective decision making
- the formal age estimation methods (encapsulated in this study by the overall estimates) tend to overage younger adults and underage middle to older aged adults (beginning around ages at death in the 50s); underageing also occurs in the subjective estimates

- the amount of bias and inaccuracy increase with age for both the overall and subjective estimates (because individual variation increases with age)
- the amount of inaccuracy, bias and standard deviation were reduced with the subjective estimates compared to the overall estimates
- the subjective estimates made it possible to give narrower age ranges for age-at-death estimates
- the age groups that were most improved by the subjective estimates in inaccuracy and bias were the youngest and oldest age groups (from age 70 and over); the improvement in the oldest age groups is particularly important, as these groups tend to be the more difficult to age
- the evidence supports the use of informal age indicators (when they are present) as additional information to provide more accurate age estimations

5.17.6 Error Testing

- interobserver error results suggest that for research questions regarding efficacy of methods, and variation in ageing rates and the expression of sexual dimorphism, comparisons using data collected by different researchers should not be attempted

5.18 Conclusion

As set out in the introduction, the main aim of this research was to gain information on the ways in which ageing and sexual dimorphism can vary in populations distinct in geographic location and time. This chapter has discussed the results of the previous chapter, including the evidence provided for population variation in ageing rates and sexual dimorphism, comparison with previous studies (where appropriate), and their implications for age and sex determination in general. It seems clear now that differences in ageing rate and sexual dimorphism may occur in populations that differ in geographic origin and/or time period, and such differences seem to be in more dimensions than previously thought. Differences can be in the rate of ageing, the shape of score/phase distributions, or timing of age-related changes and populations different in both geographic location and time period can be affected. These results fulfill the research aim of quantifying the variability in skeletal ageing and sexual dimorphism. The other research aim was to examine the relationship between the age and sex indicators and the life course. No evidence was found to link change in sexual dimorphism with age (and thus the life course), but changes in ageing rate were found across the life course.

Chapter 6: Conclusion

6.1 Introduction

In this final chapter, a general summary of the results is presented, as well as the ways in which the research addressed the original aims and the hypothesis as set out in Chapter 1. Strengths and limitations of the research are briefly discussed, followed by recommendations for future researchers that have been developed as a result of the key findings (presented in section 5.16), and suggestions for future work.

6.2 Summary of Results

6.2.1 Reliability of Methods

Results from this research support long-held notions in bioarchaeology – that the pelvis is a more reliable sex indicator than the skull, that the use of multiple indicators (sex and age) provides more reliable results, and that the pubic symphysis provides a more reliable and precise age indicator than cranial suture closure. As suggested by others (e.g. Ashley-Montagu, 1938: 372; Acsádi and Nemeskéri, 1970: 120; Key et al., 1994: 197), cranial suture closure, with its variability and wide age ranges, is best used for very broad, general age estimates, and alongside other methods of age estimation. For this sample, the "attraction of the middle" was also observed, with overageing occurring at younger ages, and underageing beginning around 50 and 60 years using the formal age estimation methods. This is similar to results found in other studies (Bedford et al., 1993: 290; Cox, 2000; Kvaal et al., 1994: 365, 367).

6.2.2 Ageing Rates and Sexual Dimorphism

Ageing rates can vary by populations separated in time and/or space, and variation in ageing rate may largely be focused on differences in one or a few specific traits under consideration for any particular ageing method (e.g. surface texture or macroporosity in the auricular surface). Ageing rates can also vary in terms of the distributions of phases or scores (that is, the shape of the phase/score distribution may vary as well as the rate), in the relative timing of an age-related characteristic, or in higher or lower occurrences of particular traits. If the reference population experienced a transition to a particular phase earlier, on average, compared to a target population, there are implications for the application of the method to other populations. The method will not perform as well, allocation accuracies will be lower (if a sex determination method), or accuracy and reliability will be lower.

Perhaps more importantly, ageing rates in a population may vary across the life course – the rate may at first be advanced compared to that of other populations, but then slow down at some point later in the life span; this may be due to alterations of relative contributions of underlying causes of degeneration (ageing) over the life course, or because of changes to the quality or intensity of one (or more) of the causes (for example, changes in activity or diet with age). This can be seen in the changing slopes of the ageing rates; where, for example, Dart Collection individuals passed through the lower pubic symphysis phases more quickly (aged at faster rates compared to Spitalfields) then 'switched' to comparatively lower rates of ageing at the higher phases.

For archaeological populations, these parameters (rates of ageing and scale of sexual dimorphism) are unknown, so error is a strong possibility. However, subjective estimates seem able to alleviate such error. It is necessary to use all possible informal age indicators, alongside all possible formal methods, to improve accuracy and precision and lessen bias. However, experience and expertise have been demonstrated to be of significance in this study, and it must be recognised that absence of any particular informal age indicator discussed in this research should not be taken as positive support for any particular age.

Some of the power from the use of "subjective" indicators (for example, osteophytes around joint margins) lies in their ability to better identify the "oldest old" in skeletal populations than the use of formal age estimation methods alone. The research presented here provides some evidence that the "oldest old" have been invisible in past studies due to the limitations of the age estimation methods, as suggested by Cox (2000: 62). However, by using all available age indicators, including the informal, subjective indicators, bioarchaeologists should be able to construct more realistic age distributions, including the presence of the older individuals in the population.

6.2.3 Use of Multiple Age Indicators

The seemingly dominant current way of combining age estimation methods to provide an age range (Garvin and Passalacqua, 2012: 430, 432) may reflect a potential solution, if researchers are willing to explicitly outline such strategies. A subjective method of weighing the most appropriate variables, using all available age or sex indicators, even if not developed into formal methods, and applying all of the user's experience and knowledge of normal human variation to provide the best possible estimate of sex and age is suggested.

Garvin and Passalacqua's (2012: 432) survey indicates that experience and expertise were already largely relied upon to narrow or expand age estimates as appropriate, to weigh methods and combine estimates based on multiple methods on an individual basis. Furthermore, there is much variation in how researchers arrive at final age estimates (Garvin and Passalacqua, 2012: 432), highlighting the subjectivity of current practice, whether acknowledged or not. As the subjective combination of methods, considering overlap of age/sex traits or indicators, on an individual and/or population level is already the general practice (Garvin and Passalacqua, 2012: 430, 432), and studies employing this practice are generally compared between researchers (out of necessity) already, it does not seem unreasonable to suggest being explicit in such practice. Extending this practice to include informal age and sex indicators is suggested, as results here indicate that the inclusion of subjective indicators increases accuracy. Furthermore, the statistical methods already developed are not well-used – Garvin and Passalacqua (2012: 432) note only 9.6% of their 145 surveyed forensic anthropologists stated that they used multifactorial (statistical) methods. Macroscopic methods are the widely-used method of choice (Falys and Lewis, 2011: 710-711), perhaps due to relative ease of use, low costs involved, and that such methods are generally fairly quick to apply.

This is not to say, however, that efforts to develop objective, replicable age and sex estimates should be halted altogether but, if statistical analyses are to be used in these efforts, researchers must be aware of the possible ways in which age and sexual dimorphism vary in the expression of indicators used. Some statistical assumptions may be violated by variation in ageing rate, the shape of phase/score distributions, or timing of particular age-related morphological changes. As noted in earlier chapters, Bayesian methods (e.g. Samworth and Gowland, 2007; Chamberlain, 2006) have also been offered as an alternative to circumvent some of these problems, but are not widely used, perhaps because of perceived complexity in using these methods.

6.3 Addressing the Aims and Hypothesis of This Research

The fundamental importance of estimates of age-at-death and sex from human skeletal remains to bioarchaeological and for forensic anthropological interpretations was stressed at the outset of this thesis. While previous studies have noted that skeletal ageing rates and sexual dimorphism vary at both the population and the individual level (e.g. Buckberry and Chamberlain, 2002; Borkan and Norris, 1980; Sherman, 1999: 11), little comprehensive work had been done to examine the range and scope of such variation. The research presented in this dissertation has addressed two aims:

1. To quantify the variability in skeletal ageing and sexual dimorphism by analysing geographically and temporally diverse skeletal populations of known age and sex and testing the efficacy of existing methods.
2. To assess the relationship between skeletal age indicators and sexual dimorphism over the life course.

It illuminated the ways in which ageing and sexual dimorphism can vary in populations separated by geography or time period, which was the first aim of this project. While it seems overly cautious to suggest that some ageing or sexing methods are inapplicable to some populations, the oft-repeated suggestion to use multiple methods certainly seems prudent. The second aim has also been fulfilled: the evidence suggests that while ageing rates can vary over the life course, no evidence for age-related trends in sexual dimorphism (after adulthood) was found here, despite other researchers' results to the contrary (e.g. Walker, 1995).

The null hypothesis tested here was that ageing rates and sexual dimorphism do not vary across human populations, and that current skeletal methods used to estimate age and sex may be used across all populations. This research has demonstrated that ageing rates and sexual dimorphism do vary across human populations; thus, the first part of the null hypothesis can be rejected. However, the evidence against worldwide application of age and sex determination methods has not been conclusive. For example, while the Suchey-Brooks method seems applicable to all populations sampled, the problem reported here of differential expression of auricular surface traits – in particular, surface texture – gives the impression that auricular surface methods should be applied to some populations with caution.

6.4 Strengths

A major strength of this research is that because only one observer collected all data, the data from different populations can be confidently compared without the problem of interobserver error, which has been reported as significant in other studies (e.g. Walrath, 2004: 136). Another strength is that data were collected from six documented collections, from four countries, enabling comparison within and between countries. Most other studies of age and sex determination methods or variability in ageing rates and sexual dimorphism sample only one population (e.g. Saunders et al., 1992; Bedford et al., 1993; Falys et al., 2006; Hens et al., 2008; Williams and Rogers, 2006).

This research has identified that scales of sexual dimorphism can vary by populations distinct in time and/or space, and that the differences do not necessarily occur in all sexually dimorphic traits for a skeletal element – it could be that only one or a few specific traits are significantly different in scale (that is, compared to the reference populations).

The scope of the variation in ageing rates is now also somewhat clearer. Variation occurs in the form of rate, shape of the phase/score distribution, and timing of age-related changes, and in higher or lower occurrences of particular traits.

6.5 Limitations

One of the limitations of the current study was the inability to tease out the varying contributions of the underlying reasons for variation in ageing rates and sexual dimorphism, such as socioeconomic status and occupation. While some of the sampled collections likely have individuals with similar socioeconomic status (for example, Pretoria and Dart), it is more difficult to reconcile the status of a sample from an earlier time period (Spitalfields) with that of the more recent collections. Furthermore, while occupations are documented for individuals from some of the collections sampled (Coimbra and Spitalfields), this information is not available for the other collections. This makes it impractical to attempt an assessment of level and intensity of physical activity that may affect ageing rates and sexual dimorphism. Other possible reasons for variation include genes, diet, lifestyle, and environment – and ruling any of these out of the equation is beyond the scope of this project.

Another limitation of this study was the relatively small intraobserver error sample; only the Grant Collection was sampled twice, as time and funding constraints made it impossible to revisit any of the other collections. It would have been preferable to test a small sample for intraobserver error from each collection. However, because the Grant Collection was the first collection visited, and data for the intraobserver error sample was collected a year later, after all other data collection, it is likely that the highest amount of error was captured.

While the value of documented collections of human skeletal remains should be stressed, the age heaping found in this sample amongst the Dart and Pretoria collections also make it clear that the reliability of documentation should be assessed prior to data collection, particularly when research involves the analysis of age-related phenomena. Measures can be taken to avoid potentially unreliable documented ages (avoiding ages ending in 0 or 5, as done here), if the threat is evaluated before sampling begins.

6.6 Recommendations

While there is still much scope for future research, some recommendations for age and sex determination may be suggested based on this research:

- If using documented collections, check for age heaping (a bar chart of the age distribution), to examine reliability of “known” ages.
- Be aware that if a particular skeletal indicator does not show the expected range of morphological changes, particularly at advanced ages, it does not necessarily mean that no individuals of old age are present; rather, that particular advanced-stage morphology may not occur in that population.

- Use multiple methods – all available skeletal indicators, including all observed informal subjective skeletal features, should be considered.
- Explicitly note if a particular feature was given more or less weight in sex or age determination due to the frequency (or paucity) of its appearance in a particular skeletal series.
- If using statistical techniques, be aware of their inherent assumptions, as differences in ageing rate or timing of age-related changes may preclude the use of some techniques.
- If objective methods (e.g. statistical) are to be used and further developed, they should not include assumptions regarding stability of rate, timing of morphological changes or phase/score distribution shape, as evidence presented here suggests that all of these may vary.

6.7 Future Research

There is still much scope for future research. While theories of senescence are still somewhat under debate, their future elucidation may also help to understand the specifics of skeletal ageing and its variation, thus allowing further improvements in methods of age estimation. The same is true for reasons behind human variation and subsequent application to differences in scale of sexual dimorphism. For example, if environmental and climatic changes are behind secular changes in craniofacial robusticity and can thus be linked to degrees of sexual dimorphism, assessments of the environment in the past may allow improved understanding of the scale of sexual dimorphism in past populations. Washburn (1949: 431) also suggested a clinal pattern of variation in sexual dimorphism, with reference to the ischio-pubis index. Additional research on the morphological indicators of sexual dimorphism to see if such a pattern of variation might exist for those traits would be interesting. Another potential avenue of study would be in research on physiologically appropriate known-age categories (rather than simply age-at-death by decade) for studies of variation in ageing rates. For instance, perhaps if bone turnover was found to slow at a particular mean age, this could be used as one such dividing point – but more research is necessary to elucidate where exactly these divisions should lie.

Further studies using documented collections from other parts of the world may also shed more light on variation in ageing and sexual dimorphism, particularly if any documented collection exists with extensive biographical information on individuals included in collections. Such research on documented collections with accurate cause of death information may also reveal the role end-of-life health has on the morphological expression of age- and sex-related indicators, although the points raised by Wood et al. (1992) must always be considered. Further exploration of intraobserver error and the role of experience in decreasing such error would also be useful. Current technologies, such as the use of geometric morphometrics and biochemical analyses as applied to age and sex estimation, also seem to hold promise for the future.

UK government-analysed statistics suggest that within 20 years, half of the adult UK population will be over 50 years old, and approximately 23% of the adult population in England will be over 65 years old (HM Government, 2008; ONS, 2012). Larger proportions of people are reaching the “oldest old” ages; in the UK, the fastest growing age group is that of people aged 80 or over (Audit Commission, 2008; Davies et al., 2012: 1793), with similar rapid growth in this group in other countries (for example, the United States: Eisdorfer et al., 1989). Numbers of people over the age of 100 are also increasing, with as many as one in four children born today expected to live beyond 100 and a projected increase of eight times the 2010 numbers of centarians by 2035 (Audit Commission, 2008; HM Government, 2008; ONS, 2011: 7). Variation between individuals, and within and between populations in terms of risk of disease, ageing rates and their interrelationship, are important to understand in terms of developing appropriate policies for older people. That ageing and risk of chronic degenerative conditions and other health conditions are heterogeneous means that policy regarding prevention of poor health and keeping older people healthy and happy must take into account the implications of such variation. While the work here concerns skeletal ageing rates, it seems likely that the parameters of skeletal ageing variation would be similar to ageing in general, and that the results of this work may be informative in terms of variation in ageing of living people. Of course, the current demographic changes mean that bioarchaeologists might benefit from paying attention to recent secular changes, as reflected in the most recent skeletal collections, as understanding variation and change occurring within our lifetimes can only benefit our interpretations of variation and change in the past.

Appendix 1: List of Documented Collections

Collection Name	Number of skeletons	Ages (years)	Males	Females	Origin	Location	Reference
Albert Szent-Gyorgyi Medical University, Department of Forensic Medicine	106	fetal and neonates				Szeged, Hungary	Usher, 2002
Amsterdam Laboratory of Anatomy and Embryology	256 crania		174+	82+	Dissection material from between AD 1883-1909; non-Jewish inhabitants of Amsterdam from across the Netherlands	Amsterdam, Netherlands	Usher's website
Banaras Hindu University	over 244	adult	over 176	over 68	Varanasi zone, India	Dept. Of Anatomy, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India	Usher, 2002
Belgian femora	over 416					Belgium	Usher, 2002
British Museum of Natural History	111 skulls				Mainly soldiers, age/nationality known; 46 recorded as insane (not no further details on type of insanity)	Natural History Museum or British Museum?	Usher, 2002

Broadbeach Osteological Collection	36		36		Ngaraangbal Aboriginal tribe's ancestral burial ground. Returned for reburial in 1985, but radiographic and photographic material available for study	Broadbeach, Australia	Usher, 2002
Brush-Bolton Collection		children				Case Western Reserve University, USA	Usher, 2002
Chiang Mai University	104	18-90	70	34	Donated remains from individuals who died at Chiang Mai University Hospital	Chiang Mai City, Thailand; kingchri@hawaii.edu	Usher, 2002
Coimbra	~500					Coimbra University, Coimbra, Portugal	Usher, 2002
Dart Collection, Dept of Anatomical Sciences	over 2000					Faculty of Health Sciences, U. of Witwatersrand, S. Africa	Usher, 2002
Dept of Anatomy and Cell Biology						Philipps University, Marburg, Germany	Usher, 2002
Dept of Anatomy, Lady Hardinge Medical College	315				Indian origin	Lady Hardinge Medical College, New Delhi, India	Usher, 2002

Drago Pervic, Institute of Anatomy Osteological Collection	300 (partially skulls)		50%	50%	Hospitals, lunatic asylums, retirement homes in Zabreb, and nearby towns; known individuals, poor Europeans	University of Zagreb Medical School, 10 000 Zagreb, Salata 3, Box 916, Croatia, ph 385 01 4566 953	Usher's website
Duckworth Osteological Collection						University of Cambridge, UK	Usher, 2002
FACES Laboratory Collection					Forensic cases	Dept of Geography and Anthropology, Louisiana State University, USA	Usher, 2002
Ferraz de Macedo						Lisbon, Portugal	Usher, 2002
Florence skull	83 (3460 skulls, 171 complete skeletons? From website)	13-62	44	39	Unclaimed indigents from Florence hospital	Museo Nazionale di Antropologia e Etnologia, Florence, Italy	Usher, 2002
Florida Atlantic University					Forensic cases	Boca Raton, FL, USA	Usher, 2002
Forensic Anthropology Data Bank	0 (?)				Forensic cases	University of Tennessee, TN, USA	Usher, 2002

Frasetto Collection	over 200	19 to over 65	over 100	over 100	Sardinian, exhumed around 1900 from Sassari cemetery	Dipartimento de Biologia Evoluzionistica Sperimentale, Bologna, Italy	Usher, 2002
Grant Collection	202	majority over 40	175	27	Unclaimed bodies	University of Toronto, Dept of Anthropology, Canada	Usher, 2002
Hamann-Todd Collection	over 3000				Cadaver	Cleveland Museum of Natural History, OH, USA	Usher, 2002
Hanged men*	3	17-26	3	0	Southern Ontario		Usher, 2002
Harvie family cemetery*	6	25-98	2	4	Southern Ontario		Usher, 2002
Highland Park Cemetery	296				Served poorhouse, used between 1826-1863		Usher, 2002
Hong Kong Collection	94	24-88	68	26	Southern China, excavation of known individuals from Wo Hop Shick cemetery	Dept of Anatomy, University of Hong Kong	Usher, 2002
Hungarian Natural History Museum, Anthropology Dept	over 10				Excavation of church basement	Natural History Museum, Budapest, Hungary	Usher, 2002

I. Gemmerich Collection	151	6 to 95	48	103	Modern cemeteries of the Vaud Canton	Dept of Anthropology and Ecology, University of Geneva, Switzerland	Usher, 2002
Institute of Anatomy	101	36 - 100	57	44	Body donors	Institute of Anatomy, University of Technology, Aachen, Germany	Usher, 2002
Institute of Forensic Medicine	large number				Cadaver donors	University of Vienna, Austria	Usher, 2002
Institute of Forensic Sciences	over 205	all	205		Modern Chinese	Ministry of Public Security, Beijing, PRC 1347 Guanfuxi Road, Shanghai, China (not sure which is correct; Shanghai is from website)	Usher, 2002
Institute of Legal Medicine	over 80	25-80	over 40	over 40	Southern Italian	Institute of Legal Medicine, University of Bari, Italy	Usher, 2002
Istituto di Anatomia (Collezione Guglielmo Romiti?)	742	all	424	317	General and mental hospital in Siena	Istituto di Anatomia, Siena, Italy	Usher, 2002
Jikei University	over 90				Japan	School of Medicine, Jikei University, Japan	Usher, 2002
School of Medicine, John Hopkins Fetal Collection						Case Western Reserve University, Cleveland, OH, USA	Usher, 2002

Medicolegal Institute at Bhopal in Central India	124		80	44	India		Usher, 2002
Mediterranean Caucasoid Collection	42 hip bones		27	15	From adult males and females of modern rural and urban population of Madrid, Spain.	Department of Morphological Sciences and Surgery, University of Alcala de Henares, Spain	Usher, 2002
M.R. Drennan Museum and Departmental Specimen Collection	~250				Cadaver	University of Cape Town, South Africa	Usher, 2002
Morphology Collection	236	30-80s			Cadavers from NYU Medical School, Long Island Medical College and the Cornell Medical School	American Museum of Natural History, New York, USA	Usher, 2002
Musee d'Anatomie Delmas-Orfila-Rouviere						V. Rene Descartes University, Paris, France	Usher, 2002
Museo do Departmento de Anatomia	492	adult			1914-1940?	Instituto de Ciencias Biomedicas da Universidade de Sao Paolo, Brazil	Usher, 2002
Museum of Pathological Anatomy	over 50 000	all			Mostly pathological	Rome, Italy	Usher, 2002

Museum Vrolik, Dept of Medicine, Academic Medical Centre	100-200				Private collection	Dept of Medicine, Academic Medical Centre, Amsterdam, Netherlands ph 31 20 566 7821	Usher, 2002
Mutter Museum, College of Physicians of Philadelphia	9		6	3	Skulls, pathological? Biographical info with skeleton (?)	19 South 22nd Street, Philadelphia, PA, 19103; ph 215-563-3737	Usher, 2002
National Museum of Health and Medicine	130+				Fetal, Civil War, forensic, pathological	Washington, DC, USA	Usher, 2002
Okamoto Research Laboratory of Dentistry						Yonago, Japan	Usher, 2002
Palmer Collection	2200	all			Pathological	Davenport, IA, USA	Usher, 2002
Quakers						School of Conservation Sciences, Bournemouth University, Talbot Campus, Poole, Dorset, England; ph 01202 595277	Usher, 2002

Royal College of Surgeons of England Museums					Odontological Museum, Hunterian Museum and the Wellcome Museums; most pathological; Odontological Museum has known age skulls colelcted by Sir Tomes	Museums, The Royal College of Surgeons of England, 35/43 Lincoln's Inn Fields, London, WC2A 3PN, UK; ph 020 7869 6570	Usher, 2002
Shellshear Museum	?				Australia, Melanesia, Oceania, Middle East (Pella collection)	Ph: (02) 9351 4529	Usher, 2002
Smithsonian Institution Fetal and Infant Collection	over 300	Fetal and infant				Smithsonian Institution, Washington, DC, USA	Usher, 2002
Spitalfields Collection		all			Christ Church, Spitalfields	Natural History Museum, London, UK	Usher, 2002
Spitalfried Hof St Johann	83	17-75	41	42		Basel, Switzerland	Usher, 2002
SR Atkinson Library of Applied Anatomy and the P&S Comparative Anatomy Collections	over 1500	*In utero* to adult			Autopsies and biological warehouses	University of the Pacific School Dentistry, San Francisco, USA	Usher, 2002
St. Bride's, London	56	22-90	26	30	Church of St. Bride, Fleet Street, London	British Museum (Natural History), London, UK	Usher, 2002

St. Thomas Anglican Church	80	all			Partial excavation of St. Thomas Anglican Church cemetery, burials 1821-1874	Belleville, Ont., Canada	Usher, 2002
State Museum of Anthropology, Dresden, Germany					Philippines, New Guinea, Malaysia, New Ireland		Usher, 2002
Stirrup Court Cemetery*	6	45-76	4	2	Southern Ontario		Usher, 2002
Suchey Pubic Collection	1225	14 to 99	739	273	Modern individuals autopsied at the Office of the Chief Medical Examiner, County of Los Angeles	California State University, Fullerton, USA	Usher, 2002
Terry Collection	over 1500				Unclaimed and donated bodies	National Museum of Natural History, Smithsonian Institution, Washington, DC, USA	Usher, 2002
Trotter Collection	over 133	Fetal				Washington State University, USA	Usher, 2002
Tubingen, Germany	over 108	4 to 86	67	41	Southwest Germany; known ages unreliable?	Institut fur Gerichtlich Medizin, Nagelestr. 5, D-72074, Tubingen, Germany	Usher, 2002

Universidad Complutense (two collections, EML 1, and EML 2)	over 132 (122; 88)	34-97 (20-91; 20-55)	over 60	over 72	Exhumed from cemetery in Madrid (contemporary Spanish, EML 1 died between 1975-1985, EML 2 died between 1941-1975)	Unidad Docente de Antropologia (Legal Medical School), Departamento de Biologia Animal I, Facultad de Biologia, Complutense University of Madrid, Dept of Biology, Spain ph (+34 1) 394 4941	Usher, 2002
University Museum, University of Tokyo	~300				Japan	University Museum, University of Tokyo; http://www.um.u-tokyo.ac.jp/en/	Usher, 2002
University of Florida Collection					Forensic cases	CA Pound Human Identification Laboratory, U. of Florida, USA	Usher, 2002
University of Indianapolis						Department of Biology, University of Indianapolis, 1400 E. Hanna Ave., Indianapolis, IN, USA 46227-3697; ph 317-788-3486	Usher, 2002

University of New Mexico, Maxwell Museum Documented Collection	~120	up to 100	77	45	Body donation program (on-going); black and white, born after 1900, well-documented with health data	Anthropology Building, Room 240, University of New Mexico, Albuquerque, USA	Usher, 2002
University of Pennsylvania Museum of Archaeology and Anthropology							Usher, 2002
University of Pretoria	over 196				South Africa (black and white)	Dept of Anatomy, U. of Pretoria, South Africa	Usher, 2002
University of Torino	1064	all	384	680	Italian, cadavers from city prisons and hospitals	Dept of Human Anatomy, Corso M. D'Azeglio 52, 10126, U. of Torino, Italy; ph +39 01167 07723	Usher, 2002
William M. Bass Donated Skeletal Collection	235	25-89			Forensic cases and donated bodies	Anthropology Dept., U. of Tennessee at Knoxville, USA	Usher, 2002
Wise family cemetery*	1	66	1	0	Southern Ontario		Usher, 2002
Wistar Institute of Anatomy and Biology	18	all	4	3	Private collection, most known age are fetal and children	Philadelphia, USA, ph 215-898-3826	Usher, 2002

Zimbabwe Museum of Human Sciences Osteological Collection	153 skulls				Black Zimbabwean population	PO Box Cy33, Harare, Zimbabwe 263-751797	Usher, 2002
Koganei Collection					Ainu skeletons	Hokkaido, Japan	Quigley, 2001
Dept of Anatomy, Sapporo College	113				Meiji Japanese skeletons	Sapporo, Japan?	Quigley, 2001
Xinjiang Medical College	115				Contemporary Chinese		Quigley, 2001
Humboldt U., School of Medicine - Anthropologische Rudolf-Virchow-Sammlung-Berlin	7000 (?)	known age??			Local cemeteries	Berlin, Germany	Quigley, 2001
'A' Series	200				1910-1930, mostly prisoners and poorhouse residents (documented)	University of Helsinki, Finland	Quigley, 2001
Tohoku Japanese crania	~200				Collected b/n 1900 and 1942	Dept of Anatomy and Anthropology, Tohoku University Medical School, Sendai, Japan	Brown and Maeda, 2004
The University of Athens Human Skeletal Reference Collection	225	0-99	114	100	214 documented; 11 undoc, mostly non-adults; exhumations.	Dept of Animal and Human Physiology, University of Athens, Greece.	Eliopoulos et al., 2007
?	No longer exists					Perth, Australia	Hunt and Albanese, 2005

?	at least 162				1900-1950, known sex.	Laboratoire d'Anthropologie biologique de l'Universite Paris 7, Paris, France.	Bruzek, 2002
	200		100	100	~1950-2008 (?) Some damage, and dried soft tissue.	Dept of Forensic Science, Medical School, University of Crete, Heraklion, Crete, Greece.	?
Rainer Collection	?					Bucharest, Romania	L'Abbe et al., 2005.
Weisbach collection	?				End of 19th C	Naturhistorisches Museum Wien, Vienna, Austria	Same as U of Vienna collection referenced in L'Abbe et al., 2005?
Vienna Collection	50 000+				Mostly pathological	Altes Allgemeines Krankenhaus (Old General Hospital), Spitalgasse 2, A-1090, Vienna, Austria	Same as above U of Vienna as in L'Abbe et al., 2005?
Military museum collection	?					Military museum, St. Petersburg, Russia	L'Abbe et al., 2005
Medico-Legal Institute of Oporto	at least 200		100	100		Porto, Portugal	De Mendonça, 2000

	at least 87				Skeletal remains excavated from cemeteries near cities of Qingdao (Shandong) and Chang Chun (Liaoning).	Yishui Medical School, Shandong, China	İşcan and Shihai, 1995
Galler Collection	~600				Pathological reference collection	National History Museum, Basel, Switzerland	Rühli et al., 2003
Cobb Laboratory collection	over 700					Cobb Laboratory, Howard University	Hunt and Albanese, 2005
Huntington Collection						National Museum of Natural History, Smithsonian Institution, Washington, DC, USA	Hunt and Albanese, 2005
(Same as below - Galloway Osteological Collection?)						Makarere College, Kampala, Uganda	Hunt and Albanese, 2005
Galloway Osteological collection	400+				Known age, sex, cause of death, date of skeletonization (1940s to 1980)	Department of Anatomy, Makerere University, Kampala, Uganda	Luboga, 2000
Stanford-Meyer Osteopathology Collection						San Diego Museum of Man, California	Albanese, 2003b

University of Iowa-Stanford Collection	100+				Individuals born mid to late 1800s (prior to major improvements in modern health care, inc advent of antibiotics and epidemiological science)	Department of Anthropology, University of Iowa	Albanese, 2003b
Museo di Storia Naturale collection	128+				Late 18th to early 20th C Italy (Florence, Siracusa, Milan)	Sezione di Antropologia ed Etnologia, Museo di Storia Naturale, University of Florence (same place as Florence skull collection?)	
Olóriz Collection	2250 skulls, documents with observations (on 15000 living, 1000 corpses)					Departamento de Anatomia y Embriologia Humana I, la Facultad de Medicina, Ciudad Universitaria, 28040 Madrid, Spain (email: secanat1@med.ucm.es)	
Anatomical Museum	~100				Known age and sex.	University of Valladolid, Spain	
University of Granada	483				Complete documented skeletons	University of Granada, Spain	Mastrangelo et al., 2011

Autonomous University of Barcelona	35				Complete skeletons from Granollers, Catalonia, 20th C, known age, sex and origin		Rissech and Steadman, 2011
Granada Osteological Collection of Identified Infants and Young Children	230	5 months gestation to 8 years old			Complete skeletons, mainly mid-20th C, from San Jose cemetery (exhumed individuals otherwise going to be incinerated or interred in communal plot).	Laboratory of Anthropology, University of Granada, Spain	Aleman et al., 2012
autopsy room collection	?				autopsy room	Department of Forensic and Insurance Medicine, Faculty of Medicine, Semmelweis University, Budapest	Wolff et al., 2012

Table A1.1. List of known documented collections of human skeletal remains

Appendix 2: Additional Results for Sex Determination Methods

A.2.1 Sexually Dimorphic Skeletal Elements: Skull

A.2.1.1 Glabella

Statistically significant differences were found, both in score distribution and median for the glabella between Grant and all other collections (see Table A2.1 for details). The differences in median between Coimbra and Pretoria neared significance. None of the other collections displayed significant differences with each other in distribution.

No age-related trends were seen in terms of percentages of correctly-sexed females or males using the glabella. The age group with the lowest proportions of correct sex identification was not consistent across collections. For instance, for Spitalfields, the lowest female and male allocation accuracies were for the 40 to 49 and 30 to 39 year age group, respectively. For Coimbra females, it was the 90 to 99 year age group, while for Lisbon males, lowest allocation accuracy was in the 50 to 59 year age group. Similarly, the age group with the highest allocation accuracy also varied by collection. For the Dart Collection, the highest female allocation accuracies were found in the 20 to 29 and 100 and over age groups, and for Lisbon females, the highest allocation accuracy was for the 20 to 29 year group. For Coimbra males, the highest allocation accuracy was in the oldest age group (90 to 99 years).

Other age groups typically had around two incorrectly-sexed individuals. Often, the oldest age category held the lowest allocation accuracy, but these also tended to be the groups that had the lowest absolute number of individuals, meaning that any wrongly-sexed individual affected the percentage correct more than they would in age groups with higher numbers of individuals. Overall, the Grant Collection had the lowest percentage of correctly-sexed females (63.6%), while the other values were fairly close to each other – the highest was for Lisbon (85.3%), but the others followed closely, with 83.3% for Pretoria, 83.1% for both Dart and Spitalfields, and 81.5% for Coimbra. In terms of overall percentage of correctly-sexed males, it was lowest for Dart and Pretoria (27.4% and 28.8%, respectively), and highest for Grant (60.8%), followed by Spitalfields (55.6%), Coimbra (53.0%) and Lisbon (44.9%).

For the sexes pooled, the significant differences were largely between Grant compared to the other collections, due to the aforementioned low number of Grant females. No significant results were found between collections when K-S and MWU tests were performed on the female-only score distributions for the glabella (Table A2.2).

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.211	.147	**.000**	**.000**	.988	.460	.111	.042	.963	.688
Dart			**.000**	**.000**	.602	.475	.840	.393	.341	.442
Grant					**.000**	**.000**	**.000**	**.000**	**.000**	**.000**
Lisbon							.221	.171	.839	.801
Pretoria									.495	.155

Table A2.1. Glabella, sexes pooled: two-sample K-S MWU test results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.877	.803	.252	1.000	.466	.986	.538	1.000	.760
Dart			.648	.281	.935	.363	.787	.414	.938	.612
Grant					.563	.127	.682	.155	.630	.206
Lisbon							1.000	.940	.999	.681
Pretoria									.997	.730

Table A2.2. Glabella, females only: two-sample K-S MWU test results

In terms of the distribution of scores, the majority of females from Spitalfields, Coimbra, Lisbon and Pretoria had scores of 1, followed by scores of 2. The opposite is true for Dart females; more had scores of 2 than 1. The Grant Collection had the highest proportion of females with scores of 3, but when the absolute numbers are observed, differences are very small (four females with scores of 3, compared to three and a half each with scores of 1 and 2). The small sample size likely affected score distribution for this collection, as only 11 females were scored for glabella morphology. Where half numbers for number of individuals are presented instead of whole numbers (e.g. 3.5), it is because individuals whose morphology did not fit one particular score (e.g. were scored 3 to 4) were divided over the score range. For all collections, few females had scores of 4, which is considered "male" morphology; Pretoria had the highest number, with three females scored as 4. Only one Dart Collection female was scored as a 5. Table A2.3, below, shows the distribution of glabella scores for females.

	Grant		Spitalfields		Coimbra		Lisbon		Dart		Pretoria	
Score	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*	*n*	% of *n*
1	3.5	31.8	27.5	44.4	31	42.5	35.5	47.3	29	39.2	35	48.6
2	3.5	31.8	24	38.7	28.5	39.0	28.5	38.0	32.5	43.9	25	34.7
3	4	36.4	9.5	15.3	12	16.4	10	13.3	9	12.2	9	12.5
4	0	0	1	1.6	1.5	2.1	1	1.3	2.5	3.4	3	4.2
5	0	0	0	0.0	0	0	0	0	1	1.4	0	0
Total	11	100.0	62	100.0	73	100.0	75	100.0	74	100.0	72	100.0

Table A2.3. Number of females with each glabella score, by collection
n: number of individuals

A.2.1.2 Supraorbital Margin

No age-related trends in female or male allocation accuracies were found using the supraorbital margin. The lowest allocation accuracies did not occur consistently in any particular age group. For example, the lowest female allocation accuracy for Spitalfields was found in the 70 to 79 year age group, while for the Grant Collection, it was in the 30 to 39 group, at 0%. The lowest male allocation accuracy for both Spitalfields and Coimbra was in the 30 to 39 year age group, but was in the 20 to 29 and 90 to 99 year age groups for Dart males. Similarly, the highest female allocation accuracies were in the 50 to 59 year age group for Spitalfields, and the 20 to 29 year age group for Coimbra. For Grant males, the highest allocation accuracy was found in the 30 to 39 year age group, while for Coimbra males, it was in the 70 to 79 year age group.

The overall percentages of correctly-sexed females ranged from 27.3% for Grant to 58.7% for Lisbon, followed closely by the Dart Collection, at 57.4%; Coimbra had 37.0% of females correctly sexed using the supraorbital margin alone, while for Spitalfields, the same value was 43.8% and for Pretoria, 46.5%. Overall, using the supraorbital margin alone, the percentage of correctly-sexed males ranges from 20.1% for the Dart Collection to a high of 50.0% for the Grant Collections. Spitalfields had a total of 38.2%, Coimbra, 40.3%, Lisbon, 25.7% and Pretoria, 29.5%.

A.2.1.3 Mastoid Process

No age-related trends in female or male allocation accuracies were found using the mastoid process. Low and high allocation accuracies occurred in all age groups. For example, the lowest percentage of correctly-sexed females for the Grant Collection were in the 20 to 29, 80 to 89 and 90 to 99 year groups, while for Spitalfields, the lowest percentage of correctly-sexed females was for the 40 to 49 year group. For Lisbon males, the lowest allocation accuracy was found in the 50 to 59 year group, while for Coimbra, the lowest allocation accuracies were in the youngest and oldest groups. Both Dart and Pretoria males had fairly low percentages across all age groups in terms of correct sex identification. Similarly, the highest allocation accuracies vary in location: for Lisbon females, the 20 to 29 and 60 to 69 year olds had the highest allocation accuracies, while for Dart, the highest allocation accuracy was in the 30 to 39 year age group. For Grant males, the highest proportion of correctly-sexed males was for the 70 to 79 year age group, while for Lisbon, it was for the 20 to 29 year age group.

The only significant differences in the mastoid process score distributions for the sexes pooled were between Grant compared to the other collections; as discussed in Chapter 5, the differences between Grant and the other collections for the sexes pooled are due to low numbers (and thus, proportions of female scores) of Grant females.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.179	.181	.014	**.000**	.844	.248	.276	.203	.500	.454
Dart			**.000**	**.000**	.436	.970	1.000	.896	.214	.621
Grant					**.001**	**.000**	**.000**	**.000**	.007	**.000**
Lisbon							.340	.959	.948	.771
Pretoria									.318	.654

Table A2.4. Mastoid process, sexes pooled: two-sample K-S and MWU test results

A.2.1.4 Nuchal Crest

As with the other morphological sex features described thus far, proportions of correctly-sexed individuals rise and fall throughout the age distribution for each collection seemingly at random using the nuchal crest. For example, the lowest allocation accuracies for females were found in the 40 to 49 year age group for Spitalfields and the 80 to 89 year age group for Coimbra. For males, the lowest allocation accuracies were found in the 60 to 69 year group for Lisbon and Pretoria, and the 40 to 49 and 20 to 29 year age groups for Dart. Similarly, the highest allocation accuracies for females were found in the 70 to 79 year group for Spitalfields, and in the 30 to 39 year group for Pretoria. For males, the highest allocation accuracies were found in the 80 to 89 year group for Coimbra, and in the 30 to 39 year age group for Lisbon.

A.2.1.5 Mental Eminence

No age-related trends were found in allocation accuracy using the mental eminence. For example, the lowest female allocation accuracies were found in the 20 to 29 year age group for Lisbon and Spitalfields, and in the 50 to 59 year group for Pretoria. For males, the lowest allocation accuracies were in the 80 to 89 year group for Spitalfields, and in the 20 to 29 and 40 to 49 year groups for Grant. The highest female allocation accuracies were similarly variable, in the 70 to 79 and 90 to 99 year group for Lisbon, and in the 20 to 29 group for Dart. For males, the highest allocation accuracies were found in the 20 to 29 and 40 to 49 year groups for Grant, and in the 60 to 69 year group for Coimbra.

No significant differences in score distribution for the mental eminence were found for either males or females alone. A few differences in median neared significance, between Spitalfields and Lisbon females and Pretoria and Spitalfields females, and between Dart and Lisbon males.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.755	.992	.999	.509	.966	.200	.984	.232	.914	.311
Dart			.995	.437	.535	.132	.487	.150	.165	.158
Grant					1.000	.927	1.000	.917	.884	.247
Lisbon							.159	.978	.326	.025
Pretoria									.296	.037

Table A2.5. Mental eminence, females only: two-sample K-S and MWU test results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.565	.089	.999	.888	.669	.432	.868	.409	.926	.675
Dart			.918	.230	.287	.013	.601	.440	.677	.256
Grant					1.000	.420	.824	.533	.989	.767
Lisbon							.231	.088	.557	.213
Pretoria									1.000	.728

Table A2.6. Mental eminence, males only: two-sample K-S and MWU test results

A.2.2 Sexually Dimorphic Skeletal Elements: Pelvis

A.2.2.1 Sciatic Notch

No age-related trends in allocation accuracy were found for the sciatic notch. For example, the lowest allocation accuracies for females were found in the youngest age groups for the Grant Collection and in the 30 to 39 year group for Pretoria, but in the 50 to 59, 70 to 79 and 80 to 89 year groups for Lisbon. For males, the lowest allocation accuracies were similarly diverse, in the 30 to 39 year group for Spitalfields and in the 90 to 99 year group for Dart. The highest female allocation accuracies were also variable by age group, in the 20 to 29 and 70 to 79 year groups for Spitalfields, and in the 90 to 99 and 50 to 59 year groups for Pretoria. For males, the highest allocation accuracies were in the 20 to 29, 80 to 89 and 90 to 99 year groups for Coimbra, and in the 20 to 29 and 100 and over group for Dart.

The significant differences found between Grant and the other collections for the sexes pooled are again due to the low number of Grant females observed (Table A2.7). The only other significant difference is in score distribution between Lisbon and Dart. For females only, no significant differences in either score distribution or central tendency were found (Table A2.8).

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.312	.168	**.001**	**.000**	.990	.599	.905	.792	.566	.397
Dart			.014	.018	.034	.055	.588	.115	.248	.057
Grant					**.000**	**.000**	**.001**	**.000**	**.000**	**.000**
Lisbon							.802	.813	.177	.659
Pretoria									.950	.601

Table A2.7. Sciatic notch, sexes pooled: two-sample K-S and MWU test results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.636	.778	.609	.368	.999	.986	.753	.201	.314	.081
Dart			.920	.680	.590	.739	.784	.394	.735	.226
Grant					.926	.370	.819	.988	.864	.717
Lisbon							.840	.205	.281	.087
Pretoria									.999	.621

Table A2.8. Sciatic notch, females only: two-sample K-S and MWU test results

A.2.2.2 Ischiopubic Ramus Ridge

No age-related trend in allocation accuracy was found for the ischiopubic ramus ridge for females or males. For example, the lowest allocation accuracies for females were found in the 30 to 39 and 90 to 99 year groups for the Grant Collection, and in the 70 to 79 year group for Coimbra. The highest allocation accuracies were found in varying age groups; for example, in the 30 to 39 and 70 to 79 year groups for Spitalfields, and in the 40 to 49 and 90 to 99 year groups for Coimbra. Similarly, the lowest male allocation accuracies were found in the 40 to 49 year group for Dart, and in the 70 to 79 year group for Pretoria, while the highest allocation accuracies were found in the 20 to 29 and 90 to 99 year groups for Coimbra, and in the 20 to 29, 70 to 79 and 100 and over age groups for Dart.

A.2.2.3 Subpubic Concavity

The allocation accuracies for the subpubic concavity showed no age-related trends. The lowest allocation accuracies for females, for example, were found in the 50 to 59 year group for Coimbra and in the 30 to 39 year group for Dart. The majority of females were sexed correctly with the subpubic concavity; accordingly, the highest allocation accuracies of 100% were found in the full range of age groups. For males, the lowest allocation accuracies were found in the 80 to 89 year group for Spitalfields, and in the 30 to 39 and 40 to 49 year groups for Coimbra, while the highest allocation accuracies range over all possible age groups (at 100%, as with the females).

Few significant differences in score distribution were found, apart from those between Grant and the other collections for the sexes pooled. As with the other traits, the subpubic concavity differences between Grant and the other collections are due to the low numbers of Grant females rather than being reflective of morphological differences in sexual dimorphism.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.070	.081	**.000**	**.000**	.430	.185	.820	.447	.997	.820
Dart			**.002**	.005	.998	.704	.769	.351	.018	.071
Grant					**.003**	**.001**	**.000**	**.000**	**.000**	**.000**
Lisbon							1.000	.577	.148	.153
Pretoria									.365	.383

Table A2.9. Subpubic concavity, sexes pooled: two-sample K-S and MWU test results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.996	.240	.948	.026	.864	.025	1.000	.676	.934	.040
Dart			.999	.257	.999	.269	.995	.435	1.000	.357
Grant					1.000	.766	.979	.057	1.000	.633
Lisbon							.948	.055	1.000	.844
Pretoria									.983	.084

Table A2.10. Subpubic concavity, females only: two-sample K-S and MWU test results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.301	.005	.997	.346	.444	.015	.747	.086	1.000	.459
Dart			.932	.075	1.000	.799	1.000	.295	.141	**.001**
Grant					.987	.138	1.000	.469	.814	.126
Lisbon							1.000	.437	.220	**.003**
Pretoria									.413	.026

Table A2.11. Subpubic concavity, males only: two-sample K-S and MWU test results

A.2.2.4 Ventral Arc

No age-related trends in allocation accuracies were found for the ventral arc. The highest allocation accuracies occurred in all age groups. As noted in Chapter 5, while the youngest and oldest groups had lower allocation accuracies, these groups also had lower absolute numbers of females; throughout the age groups, only one or two females tended to be incorrectly identified, regardless of each female's contribution to the percentage of correct sex identifications to the age group. The ventral arc performed well for males as well as females. The lowest allocation accuracies for males were found in various age groups. For example, in the 20 to 29 and 60 to 69 year group for Grant, and in the 80 to 89 and 40 to 49 year group for Spitalfields. Many age groups had allocation accuracies of 100%.

No significant differences in score distribution were found for the ventral arc, apart from those between Grant and the other collections for the sexes pooled. As before, the ventral arc differences between Grant and the other collections are because of the low numbers of Grant females.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.926	.752	**.002**	**.000**	1.000	.973	.994	.807	.808	.329
Dart			**.000**	**.000**	1.000	.796	1.000	.951	.451	.500
Grant					**.001**	**.000**	**.000**	**.000**	**.000**	**.000**
Lisbon							1.000	.843	.529	.351
Pretoria									.546	.457

Table A2.12. Ventral arc, sexes pooled: two-sample K-S and MWU test results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.686	.036	1.000	.600	.996	.294	.559	.012	1.000	.657
Dart			.766	.029	1.000	.309	1.000	.681	.944	.105
Grant					.978	.206	.686	.010	1.000	.416
Lisbon							.999	.148	1.000	.581
Pretoria									.866	.041

Table A2.13. Ventral arc, females only: two-sample K-S and MWU test results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.966	.377	1.000	.737	1.000	.699	1.000	.571	.999	.214
Dart			1.000	.601	.996	.211	.989	.151	.997	.657
Grant					1.000	.464	1.000	.363	1.000	.347
Lisbon							1.000	.882	.981	.107
Pretoria									.962	.066

Table A2.14. Ventral arc, males only: two-sample K-S and MWU test results

A.2.3 Metrical Method

Albanese's (2003a) metrical method was tested with independent t-tests to look for sex differences for each measurement within each collection (Table A2.15). Significant differences were found in the majority of cases. Only iliac breadth and SPRL had non-significant differences for Dart, Lisbon, Pretoria and Spitalfields (iliac breadth) and for Coimbra and Grant (SPRL).

However, as noted in Chapter 4, non-significant differences do not necessarily mean that the distribution of the measurement is not bimodal or incapable of discriminating sex.

	Femur Max Length	Femur Max Diameter	Femur Epicondylar Breadth	Hip Bone Height	Iliac Breadth	SPRL	AIL
Coimbra	**.000**	**.000**	**.000**	**.000**	**.035**	.126	**.000**
Dart	**.000**	**.000**	**.000**	**.000**	.105	**.002**	**.000**
Grant	**.000**	**.000**	**.000**	**.000**	**.000**	.110	**.000**
Lisbon	**.000**	**.000**	**.000**	**.000**	.305	**.000**	**.000**
Pretoria	**.000**	**.000**	**.000**	**.000**	.223	**.000**	**.000**
Spitalfields	**.000**	**.000**	**.000**	**.000**	.053	**.015**	**.000**

Table A2.15. Sex differences in measurements used in Albanese's metric method

A.2.4 Allocation Accuracies of Pelvic and Skull Traits by Sex for Each Age Group

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	0			6	6	100.0	10	9.5	95.0	10	10	100.0	9	9	100.0	9	9	100.0
30-39	1	1	100.0	9	8.5	94.4	10	9	90.0	9	8.5	94.4	10	8.5	85.0	11	9.5	86.4
40-49	0			9	6	66.7	10	9	90.0	10	8.5	85.0	9	8.5	94.4	10	9	90.0
50-59	0			8	6.5	81.3	10	7.5	75.0	10	9.5	95.0	9	8	88.9	11	9.5	86.4
60-69	4	2	50.0	10	7	70.0	10	7.5	75.0	9	6	66.7	10	7.5	75.0	10	7	70.0
70-79	3	2	66.7	10	8	80.0	10	7.5	75.0	10	8.5	85.0	10	6.5	65.0	9	8.5	94.4
80-89	1	1	100.0	10	9.5	95.0	10	7.5	75.0	10	8	80.0	10	8	80.0	10	6.5	65.0
90-99	2	1	50.0		0		3	2	66.7	7	5	71.4	6	4.5	75.0	2	1	50.0
100+	0				0		0			0			1	1	100.0	0		
Total	11	7	63.6	62	51.5	83.1	73	59.5	81.5	75	64	85.3	74	61.5	83.1	72	60	83.3

Table A2.16. Glabella, number and percentage of correctly-sexed females
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	2	0.5	25.0	5	2	40.0	9	3	33.3	10	4.5	45.0	10	2.5	25.0	10	2	20.0
30-39	9	6	66.7	7	2.5	35.7	10	4	40.0	10	4.5	45.0	10	2.5	25.0	10	3	30.0
40-49	9	6.5	72.2	10	4.5	45.0	10	5.5	55.0	9	3.5	38.9	10	3.5	35.0	10	3.5	35.0
50-59	11	7	63.6	9	7	77.8	10	7.5	75.0	10	3.5	35.0	10	3.5	35.0	10	2	20.0
60-69	8	6.5	81.3	9	5	55.6	11	6	54.5	10	5.5	55.0	10	2	20.0	10	2.5	25.0
70-79	12	5	41.7	9	6	66.7	10	7	70.0	9	4	44.4	10	2.5	25.0	9	4.5	50.0
80-89	9	5	55.6	4	2	50.0	6	1.5	25.0	10	5	50.0	10	4	40.0	10	2.5	25.0
90-99				1	1	100.0	1	1	100.0	0			9	1.5	16.7	4	1	25.0
100+					0		0			0			3	0.5	16.7	0		
Total	60	36.5	60.8	54	30	55.6	67	35.5	53.0	68	30.5	44.9	82	22.5	27.4	73	21	28.8

Table A2.17. Glabella, number and percentage of correctly-sexed males
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	0			8	4.5	56.3	10	6	60.0	10	5.5	55.0	9	7.5	83.3	9	6	66.7
30-39	1	0	0.0	9	4	44.4	10	4	40.0	9	6.5	72.2	10	6.5	65.0	11	6.5	59.1
40-49	0			10	3	30.0	10	3.5	35.0	10	6	60.0	9	4.5	50.0	10	3	30.0
50-59	0			8	5.5	68.8	10	4.5	45.0	10	7	70.0	9	5	55.6	11	6.5	59.1
60-69	4	0.5	12.5	10	5.5	55.0	10	1.5	15.0	9	3.5	38.9	10	6	60.0	10	6	60.0
70-79	3	1.5	50.0	9	2.5	27.8	10	2.5	25.0	10	5.5	55.0	10	3.5	35.0	9	3.5	38.9
80-89	1	1	100.0	10	3	30.0	10	5	50.0	10	7	70.0	10	6	60.0	10	1	10.0
90-99	2				0		3	0	0.0	7	3	42.9	6	2.5	41.7	2	1	50.0
100+	0				0		0			0			1	1	100.0	0		
Total	11	3	27.3	64	28	43.8	73	27	37.0	75	44	58.7	74	42.5	57.4	72	33.5	46.5

Table A2.18. Supraorbital margin, number and percentage of correctly-sexed females
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	n Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	2	1	50.0	5	3	60.0	9	3.5	38.9	10	2	20.0	10	0	0.0	10	5	50.0
30-39	9	7	77.8	8	1.5	18.8	10	2.5	25.0	10	1	10.0	10	1.5	15.0	10	1	10.0
40-49	9	5	55.6	9	4.5	50.0	10	5	50.0	10	4.5	45.0	10	4	40.0	10	3.5	35.0
50-59	11	4	36.4	9	2.5	27.8	10	3	30.0	10	2	20.0	10	3	30.0	10	5	50.0
60-69	9	1	11.1	9	2	22.2	11	4	36.4	10	2.5	25.0	10	3.5	35.0	10	2	20.0
70-79	12	6	50.0	9	3.5	38.9	10	5.5	55.0	10	3.5	35.0	10	2	20.0	9	2	22.2
80-89	9	6.5	72.2	5	3	60.0	6	3	50.0	10	2.5	25.0	10	2	20.0	10	2	20.0
90-99				1	1	100.0	1	0.5	50.0	0			9	0	0.0	4	1	25.0
100+					0		0			0			3	0.5	16.7	0		
Total	61	30.5	50.0	55	21	38.2	67	27	40.3	70	18	25.7	82	16.5	20.1	73	21.5	29.5

Table A2.19. Supraorbital margin, number and percentage of correctly-sexed males
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	1	0	0.0	9	5	55.6	9	7	77.8	10	9	90.0	9	5	55.6	9	5	55.6
30-39	0			10	8.5	85.0	10	4	40.0	9	7	77.8	10	8	80.0	11	7.5	68.2
40-49	0			8	3.5	43.8	10	5.5	55.0	10	7.5	75.0	9	6	66.7	10	6.5	65.0
50-59	0			8	4	50.0	10	4.5	45.0	10	6.5	65.0	9	4.5	50.0	11	3	27.3
60-69	4	2.5	62.5	10	5.5	55.0	10	5.5	55.0	10	9	90.0	10	7.5	75.0	10	6.5	65.0
70-79	3	1.5	50.0	9	6	66.7	10	10	100.0	10	6.5	65.0	10	3.5	35.0	9	6	66.7
80-89	1		0.0	10	5	50.0	10	5	50.0	10	5	50.0	10	5	50.0	10	4.5	45.0
90-99	2		0.0		0		3	0.5	16.7	7	6	85.7	6	3.5	58.3	2	1	50.0
100+	0				0		0			0			1	0	0.0	0		
Total	11	4	36.4	64	37.5	58.6	72	42	58.3	76	56.5	74.3	74	43	58.1	72	40	55.6

Table A2.20. Mastoid process, number and percentage of correctly-sexed females

n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	2	0	0.0	6	2.5	41.7	9	2	22.2	10	7.5	75.0	10	1.5	15.0	10	1	10.0
30-39	8	3	37.5	7	2.5	35.7	10	4	40.0	10	5.5	55.0	10	2.5	25.0	10	2	20.0
40-49	9	6.5	72.2	7	1	14.3	10	4	40.0	10	4.5	45.0	10	3.5	35.0	10	3	30.0
50-59	9	4	44.4	8	6.5	81.3	10	8	80.0	10	3	30.0	10	1.5	15.0	10	4.5	45.0
60-69	7	3.5	50.0	11	5.5	50.0	11	5.5	50.0	10	5	50.0	10	3.5	35.0	10	2.5	25.0
70-79	12	10	83.3	11	5.5	50.0	10	6	60.0	9	6	66.7	10	2.5	25.0	9	2.5	27.8
80-89	9	6	66.7	4	2	50.0	6	5	83.3	10	5.5	55.0	10	4	40.0	10	2.5	25.0
90-99				1	1	100.0	1	0	0.0	0			9	4	44.4	4	1	25.0
100+					0		0			0			3	0	0.0	0		
Total	56	33	58.9	55	26.5	48.2	67	34.5	51.5	69	37	53.6	82	23	28.0	73	19	26.0

Table A2.21. Mastoid process, number and percentage of correctly-sexed males
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	1	1	100.0	8	6	75.0	10	10	100.0	9	6	66.7	9	9	100.0	9	6	66.7
30-39	1	0	0.0	9	6	66.7	10	7	70.0	9	4	44.4	10	6	60.0	11	6	54.5
40-49	0			7	3	42.9	10	8	80.0	10	4	40.0	9	7	77.8	10	9	90.0
50-59	0			8	5	62.5	10	6.5	65.0	10	5	50.0	9	8	88.9	11	2	18.2
60-69	4	1	25.0	8	5	62.5	10	8	80.0	10	8	80.0	10	7	70.0	10	4	40.0
70-79	3	2.5	83.3	9	8.5	94.4	10	7.5	75.0	10	5	50.0	10	6	60.0	9	4	44.4
80-89	1	1	100.0	10	9	90.0	10	6	60.0	10	6	60.0	10	5	50.0	10	2	20.0
90-99	2	0.5	25.0		0		3	2	66.7	7	4	57.1	6	1	16.7	2	2	100.0
100+	0				0		0			0			1	1	100.0	0		
Total	12	6	50.0	59	42.5	72.0	73	55	75.3	75	42	56.0	74	50	67.6	72	35	48.6

Table A2.22. Nuchal crest, number and percentage of correctly-sexed females
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	2	2	100.0	6	3	50.0	9	2.5	27.8	10	5.5	55.0	10	1.5	15.0	10	3	30.0
30-39	9	7.5	83.3	7	3	42.9	10	4	40.0	10	7.5	75.0	10	4	40.0	10	4.5	45.0
40-49	9	5.5	61.1	7	1	14.3	10	3.5	35.0	10	6	60.0	10	3	30.0	10	3	30.0
50-59	11	5.5	50.0	8	6.5	81.3	10	2	20.0	10	7	70.0	10	1	10.0	10	2	20.0
60-69	9	5.5	61.1	9	3	33.3	11	7	63.6	10	3.5	35.0	10	3	30.0	10	1	10.0
70-79	12	6	50.0	9	6	66.7	10	3.5	35.0	10	6.5	65.0	10	2.5	25.0	9	5	55.6
80-89	9	8.5	94.4	3	1.5	50.0	6	4	66.7	10	6.5	65.0	10	4	40.0	10	7	70.0
90-99				1	0	0.0	1	0	0.0	0			9	4.5	50.0	4	2.5	62.5
100+					0		0			0			3	1	33.3	0		
Total	61	40.5	66.4	50	24	48.0	67	26.5	39.6	70	42.5	60.7	82	24.5	29.9	73	28	38.4

Table A2.23. Nuchal crest, number and percentage of correctly-sexed males
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	0			8	3	37.5	10	7	70.0	9	3.5	38.9	10	10	100.0	9	6	66.7
30-39	0			9	6.5	72.2	10	5.5	55.0	9	6	66.7	10	6.5	65.0	8	5.5	68.8
40-49	1	0.5	50.0	7	6	85.7	10	6	60.0	10	5.5	55.0	9	5.5	61.1	7	5	71.4
50-59	0			8	6.5	81.3	10	7	70.0	10	6	60.0	9	6.5	72.2	9	4	44.4
60-69	2	2	100.0	10	7	70.0	10	8	80.0	10	6.5	65.0	10	9	90.0	10	8.5	85.0
70-79	1	1	100.0	9	7	77.8	10	8.5	85.0	9	8	88.9	10	6.5	65.0	9	5	55.6
80-89	1	0	0.0	9	9	100.0	10	8.5	85.0	9	6	66.7	10	8	80.0	9	5	55.6
90-99	2	1	50.0				3	1.5	50.0	7	6	85.7	6	5	83.3	2	1.5	75.0
100+	0						0			0			1	1	100.0	0		
Total	7	4.5	64.3	60	45	75.0	73	52	71.2	73	47.5	65.1	75	58	77.3	63	40.5	64.3

Table A2.24. Mental eminence, number and percentage of correctly-sexed females
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	1	0	0.0	4	0.5	12.5	9	1.5	16.7	10	3.5	35.0	9	1.5	16.7	8	1	12.5
30-39	4	1.5	37.5	9	3	33.3	10	2.5	25.0	9	4	44.4	10	2	20.0	10	3.5	35.0
40-49	3	0	0.0	7	3.5	50.0	10	2	20.0	10	4	40.0	10	1.5	15.0	10	3	30.0
50-59	6	2.5	41.7	9	1	11.1	10	4	40.0	10	6	60.0	10	3.5	35.0	8	4	50.0
60-69	6	1	16.7	11	3.5	31.8	11	5.5	50.0	10	3	30.0	10	4.5	45.0	10	3.5	35.0
70-79	10	3.5	35.0	8	4	50.0	10	4	40.0	10	3	30.0	10	2.5	25.0	7	2.5	35.7
80-89	5	3.5	70.0	4	0	0.0	6	1.5	25.0	10	3.5	35.0	10	3	30.0	9	1	11.1
90-99		0		1	1	100.0	1	0	0.0	0			8	0	0.0	3	0.5	16.7
100+		0			0		0			0			3	0.5	16.7	0		
Total	35	12	34.3	53	16.5	31.1	67	21	31.3	69	27	39.1	80	19	23.8	65	19	29.2

Table A2.25. Mental eminence, number and percentage of correctly-sexed males
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	2	0	0.0	8	7	87.5	10	6.5	65.0	10	5	50.0	10	4	40.0	9	4.5	50.0
30-39	1	0	0.0	10	7	70.0	10	5.5	55.0	9	6.5	72.2	9	6	66.7	10	4	40.0
40-49	2	2	100.0	9	6.5	72.2	10	5.5	55.0	10	8	80.0	9	5	55.6	10	6.5	65.0
50-59				10	3.5	35.0	10	6	60.0	10	4	40.0	10	3.5	35.0	10	4.5	45.0
60-69	5	5	100.0	10	5	50.0	10	3.5	35.0	10	4.5	45.0	8	6.5	81.3	11	8.5	77.3
70-79	5	3	60.0	10	7.5	75.0	10	4.5	45.0	10	4	40.0	10	3	30.0	10	7	70.0
80-89	1	0.5	50.0	8	3.5	43.8	10	4.5	45.0	10	4	40.0	10	6.5	65.0	10	4.5	45.0
90-99	2	1	50.0	0			3	1.5	50.0	7	4.5	64.3	6	3.5	58.3	2	2	100.0
100+				0			0			0			1	0	0.0	0		
Total	18	11.5	63.9	65	40	61.5	73	37.5	51.4	76	40.5	53.3	73	38	52.1	72	41.5	57.6

Table A2.26. Sciatic notch, number and percentage of correctly-sexed females
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	3	1.5	50.0	6	4	66.7	9	9	100.0	10	8	80.0	10	7	70.0	10	6	60.0
30-39	10	6.5	65.0	9	5	55.6	10	6.5	65.0	10	7	70.0	10	7	70.0	10	8	80.0
40-49	9	7	77.8	10	6	60.0	10	8	80.0	10	7.5	75.0	10	7	70.0	10	8.5	85.0
50-59	11	7.5	68.2	9	8	88.9	10	8	80.0	10	7.5	75.0	10	9	90.0	10	9	90.0
60-69	10	8	80.0	9	7	77.8	11	8	72.7	10	6.5	65.0	10	10	100.0	10	9.5	95.0
70-79	13	11.5	88.5	9	7.5	83.3	10	8	80.0	10	8.5	85.0	10	8	80.0	10	9	90.0
80-89	9	8.5	94.4	3	3	100.0	6	6	100.0	10	6	60.0	10	8.5	85.0	10	5	50.0
90-99				1	1	100.0	1	1	100.0	0			9	4.5	50.0	4	3	75.0
100+					0		0			0			3	3	100.0	0		
Total	65	50.5	77.7	56	41.5	74.1	67	54.5	81.3	70	51	72.9	82	64	78.0	74	58	78.4

Table A2.27. Sciatic notch, number and percentage of correctly-sexed males

n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	1	1	100.0	4	2	50.0	8	4	50.0	8	5	62.5	9	7	77.8	9	8	88.9
30-39	1	0	0.0	8	8	100.0	10	6	60.0	8	8	100.0	9	7	77.8	10	9	90.0
40-49	2	1	50.0	9	7	77.8	9	8	88.9	9	8	88.9	9	6	66.7	10	10	100.0
50-59				8	7	87.5	10	8	80.0	10	9	90.0	10	6	60.0	10	9	90.0
60-69	5	5	100.0	9	7	77.8	10	8	80.	7	6	85.7	8	7	87.5	10	10	100.0
70-79	5	4	80.0	6	6	100.0	10	4	40.0	7	5	71.4	10	8	80.0	10	8	80.0
80-89	1	1	100.0	10	7	70.0	10	7	70.0	10	8	80.0	10	5	50.0	10	9	90.0
90-99	1		0.0	0			3	3	100.0	5	2	40.0	6	4	66.7	2	2	100.0
100+				0			0			0			1	1	100.0	0		
Total	16	12	75.0	54	44	81.5	70	48	68.6	64	51	79.7	72	51	70.8	71	65	91.5

Table A2.28. Ischiopubic ramus ridge, number and percentage of correctly-sexed females
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	3	2	66.7	4	3	75.0	8	8	100.0	10	9	90.0	10	10	100.0	10	9	90.0
30-39	10	9	90.0	8	5	62.5	9	7	77.8	10	6	60.0	10	9	90.0	10	7	70.0
40-49	9	9	100.0	9	8	88.9	10	9	90.0	10	8	80.0	10	5	50.0	10	7	70.0
50-59	11	10	90.9	7	6	85.7	10	8	80.0	10	9	90.0	10	8	80.0	10	6	60.0
60-69	10	10	100.0	7	5	71.4	11	9	81.8	9	5	55.6	10	7	70.0	10	7	70.0
70-79	13	13	100.0	6	6	100.0	10	9	90.0	7	7	100.0	10	10	100.0	10	4	40.0
80-89	9	9	100.0	2	2	100.0	6	4	66.7	10	8	80.0	10	8	80.0	10	7	70.0
90-99				1	1	100.0	1	1	100.0	0			9	8	88.9	4	3	75.0
100+							0			0			3	3	100.0	0		
Total	65	62	95.4	44	36	81.8	65	55	84.6	66	52	78.8	82	68	82.9	74	50	67.6

Table A2.29. Ischiopubic ramus ridge, number and percentage of correctly-sexed males
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	2	2	100.0	6	6	100.0	9	9	100.0	9	8	88.9	9	9	100.0	9	9	100.0
30-39	1	1	100.0	8	8	100.0	10	10	100.0	9	9	100.0	9	7	77.8	10	10	100.0
40-49	2	2	100.0	9	9	100.0	9	9	100.0	9	8	88.9	9	9	100.0	10	10	100.0
50-59				8	8	100.0	10	9	90.0	10	9	90.0	10	9	90.0	10	10	100.0
60-69	5	5	100.0	9	8	88.9	10	10	100.0	9	7	77.8	8	8	100.0	10	10	100.0
70-79	5	5	100.0	6	6	100.0	10	10	100.0	7	7	100.0	10	8	80.0	10	10	100.0
80-89	1	1	100.0	10	10	100.0	10	10	100.0	10	10	100.0	10	10	100.0	10	10	100.0
90-99	2	1	50.0	0			3	3	100.0	6	6	100.0	6	6	100.0	2	2	100.0
100+				0			0			0			1	1	100.0	0		
Total	18	17	94.4	56	55	98.2	71	70	98.6	69	64	92.8	72	67	93.1	71	71	100.0

Table A2.30. Subpubic concavity, number and percentage of correctly-sexed females
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	3	3	100.0	4	4	100.0	9	9	100.0	10	10	100.0	10	10	100.0	10	9	90.0
30-39	10	9	90.0	9	9	100.0	10	9	90.0	10	10	100.0	10	8	80.0	10	9	90.0
40-49	9	9	100.0	9	8	88.9	10	9	90.0	10	10	100.0	10	9	90.0	10	9	90.0
50-59	11	10	90.9	8	7	87.5	10	10	100.0	10	10	100.0	10	10	100.0	10	10	100.0
60-69	10	8	80.0	7	6	85.7	11	11	100.0	9	8	88.9	10	10	100.0	10	10	100.0
70-79	13	13	100.0	6	6	100.0	10	10	100.0	8	8	100.0	10	10	100.0	10	10	100.0
80-89	9	9	100.0	2	1	50.0	6	6	100.0	10	9	90.0	10	10	100.0	10	8	80.0
90-99				1	1	100.0	1	1	100.0	0			9	8	88.9	4	4	100.0
100+							0			0			3	3	100.0	0		
Total	65	61	93.8	46	42	91.3	67	65	97.0	67	65	97.0	82	78	95.1	74	69	93.2

Table A2.31. Subpubic concavity, number and percentage of correctly-sexed males
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	1	1	100.0	4	3	75.0	7	5	71.4	9	7	77.8	9	9	100.0	9	9	100.0
30-39	1	1	100.0	8	8	100.0	10	9	90.0	8	8	100.0	9	7	77.8	10	10	100.0
40-49	2	2	100.0	9	9	100.0	9	9	100.0	9	8	88.9	9	9	100.0	10	10	100.0
50-59				8	7	87.5	10	9	90.0	10	8	80.0	10	9	90.0	10	10	100.0
60-69	5	5	100.0	9	9	100.0	10	10	100.0	7	6	85.7	8	8	100.0	10	10	100.0
70-79	5	4	80.0	6	4	66.7	10	10	100.0	7	7	100.0	10	9	90.0	10	10	100.0
80-89	1	1	100.0	9	9	100.0	9	8	88.9	10	10	100.0	10	10	100.0	10	8	80.0
90-99	2	1	50.0	0			3	2	66.7	6	6	100.0	6	6	100.0	2	2	100.0
100+				0			0			0			1	1	100.0	0		
Total	17	15	88.2	53	49	92.5	68	62	91.2	66	60	90.9	72	68	94.4	71	69	97.2

Table A2.32. Ventral arc, number and percentage of correctly-sexed females
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
	n	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr	*n*	*n* Corr	% Corr
20-29	3	2	66.7	3	3	100.0	8	8	100.0	10	10	100.0	10	9	90.0	10	10	100.0
30-39	10	10	100.0	8	8	100.0	9	9	100.0	10	10	100.0	10	8	80.0	10	9	90.0
40-49	9	9	100.0	9	8	88.9	9	8	88.9	10	10	100.0	10	9	90.0	10	10	100.0
50-59	11	10	90.9	8	8	100.0	10	10	100.0	10	9	90.0	10	9	90.0	10	10	100.0
60-69	10	9	90.0	7	7	100.0	11	10	90.9	9	8	88.9	9	8	88.9	10	10	100.0
70-79	13	13	100.0	6	6	100.0	10	9	90.0	7	7	100.0	10	10	100.0	10	10	100.0
80-89	9	9	100.0	2	1	50.0	6	6	100.0	10	9	90.0	10	9	90.0	10	9	90.0
90-99				1	1	100.0	1	1	100.0	0			9	8	88.9	4	4	100.0
100+							0			0			3	3	100.0	0		
Total	65	62	95.4	44	42	95.5	64	61	95.3	66	63	95.5	81	73	90.1	74	72	97.3

Table A2.33. Ventral arc, number and percentage of correctly-sexed males
n: number of individuals; *n* Corr: number of individuals sexed correctly; % Corr: percentage of individuals sexed correctly

Appendix 3: Allocation Accuracies for Pelvis, Skull and Pelvis and Skull Combined by Age Group

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
Age Group	*n*	*n* corr	% corr	*n*	*n* corr	% corr	*n*	*n* corr	% corr	*n*	*n* corr	% corr	*n*	*n* corr	% corr	*n*	*n* corr	% corr
20-29																		
F	2	2	100	8	8	100	10	10	100	10	9	90	10	9	90	9	9	100
M	3	3	100	6	6	100	9	9	100	10	10	100	10	10	100	10	10	100
P	5	5	100	14	14	100	19	19	100	20	19	95	20	19	95	19	19	100
30-39																		
F	1	1	100	8	8	100	10	10	100	9	9	100	9	7	78	10	10	100
M	10	10	100	9	8	89	10	9	90	10	10	100	10	10	100	10	9	90
P	11	11	100	17	16	94	20	19	95	19	19	100	19	17	89	20	19	95
40-49																		
F	2	2	100	10	10	100	10	9	90	10	9	90	9	9	100	10	10	100
M	9	9	100	10	10	100	10	9	90	10	10	100	10	9	90	10	10	100
P	11	11	100	20	20	100	20	18	90	20	19	95	19	18	95	20	20	100
50-59																		
F	0			10	10	100	10	9	90	10	9	90	10	9	90	10	10	100
M	11	10	91	9	9	100	10	10	100	10	10	100	10	10	100	10	10	100
P				19	19	100	20	19	95	20	19	95	20	19	95	20	20	100
60-69																		
F	5	5	100	10	9	90	10	10	100	10	8	80	8	8	100	11	10	91
M	10	9	90	8	8	100	11	11	100	10	9	90	10	10	100	10	10	100
P	15	14	93	18	17	94	21	21	100	20	17	85	18	18	100	21	20	95
70-79																		
F	5	5	100	10	10	100	10	10	100	10	10	100	10	8	80	10	10	100
M	13	13	100	10	9	90	10	10	100	10	10	100	10	10	100	10	10	100
P	18	18	100	20	19	95	20	20	100	20	20	100	20	18	90	20	20	100
80-89																		

F	1	1	100	10	10	100	10	10	100	10	10	100	10	10	100	10	9	90
M	9	9	100	3	2	67	6	6	100	10	9	90	10	10	100	10	9	90
P	10	10	100	13	12	92	16	16	100	20	19	95	20	20	100	20	18	90
90-99																		
F	2	1	50	0			3	3	100	7	7	100	6	6	100	2	2	100
M	0			1	1	100	1	1	100	0			9	8	89	4	4	100
P							4	4	100				15	14	93	6	6	100
100+																		
F	0			0			0			0			1	1	100	0		
M	0			0			0			0			3	3	100	0		
P	0			0			0						4	4	100			
Total	83	80	96.4	122	118	96.7	140	136	97.1	146	139	95.2	135	129	95.6	146	142	97.3

Table A3.1. Pelvis sex assessment; numbers and percentages correct by age group and sex for each collection

F = female; M = male; P = sexes pooled; *n* = number of individuals; *n* corr = number of correct sex assignments; % corr = percentage correct sex assignments.

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
Age Group	*n*	*n* corr	% corr	*n*	*n* corr	% corr	*n*	*n* corr	% corr	*n*	*n* corr	% corr	*n*	*n* corr	% corr	*n*	*n* corr	% corr
20-29																		
F	1	1	100	9	8	89	10	10	100	10	10	100	10	10	100	9	8	89
M	2	2	100	6	5	83	9	5	56	10	8	80	10	3	30	10	5	50
P	3	3	100	15	13	87	19	15	79	20	18	90	20	13	65	19	13	68
30-39																		
F	1	1	100	10	8	80	10	10	100	9	9	100	10	9	90	11	10	91
M	9	9	100	11	6	55	10	8	80	10	8	80	10	5	50	10	4	40
P	10	10	100	21	14	67	20	18	90	19	17	89	20	14	70	21	14	67
40-49																		
F	1	1	100	10	7	70	10	9	90	10	9	90	9	9	100	10	8	80

M	9	9	100	10	8	80	10	9	90	10	8	80	10	6	60	10	5	50
P	10	10	100	20	15	75	20	18	90	20	17	85	19	15	79	20	13	65
50-59																		
F	0			8	7	88	10	9	90	10	9	90	9	8	89	11	5	45
M	11	9	82	10	10	100	10	8	80	10	8	80	10	8	80	10	6	60
P				18	17	94	20	17	85	20	17	85	19	16	84	21	11	52
60-69																		
F	4	3	75	10	10	100	10	8	80	10	8	80	10	10	100	10	9	90
M	9	8	89	11	7	64	11	9	82	10	8	80	10	3	30	10	3	30
P	13	11	85	21	17	81	21	17	81	20	16	80	20	13	65	20	12	60
70-79																		
F	3	3	100	9	9	100	10	10	100	10	9	90	10	7	70	9	6	67
M	12	11	92	11	8	73	10	9	90	10	8	80	10	4	40	9	5	56
P	15	14	93	20	17	85	20	19	95	20	17	85	20	11	55	18	11	61
80-89																		
F	1	1	100	10	10	100	10	7	70	10	9	90	10	9	90	10	3	30
M	9	9	100	5	3	60	6	5	83	10	9	90	10	5	50	10	6	60
P	10	10	100	15	13	87	16	12	75	20	18	90	20	14	70	20	9	45
90-99																		
F	2	2	100	0			3	2	67	7	7	100	6	6	100	2	2	100
M	0			1	1	100	1	1	100	0			9	4	44	4	2	50
P							4	3	75				15	10	67	6	4	67
100+																		
F	0			0			0			0			1	1	100	0		
M	0			0			0			0			3	2	67	0		
P	0			0			0						4	3	75			
Total	74	69	93.24	131	107	81.68	140	119	85.00	146	127	86.99	157	109	69.43	145	87	60.00

Table A3.2. Skull sex assessment; numbers and percentages correct by age group and sex for each collection

F = female; M = male; P = sexes pooled; *n* = number of individuals; *n* corr = number of correct sex assignments; % corr = percentage correct sex assignments.

	Grant			Spitalfields			Coimbra			Lisbon			Dart			Pretoria		
Age Group	*n*	*n* corr	% corr	*n*	*n* corr	% corr	*n*	*n* corr	% corr	*n*	*n* corr	% corr	*n*	*n* corr	% corr	*n*	*n* corr	% corr
20-29																		
F	1	1	100.00	8	8	100.00	10	10	100.00	10	9	90.00	10	10	100.00	9	9	100.00
M	2	2	100.00	6	6	100.00	9	9	100.00	10	10	100.00	10	10	100.00	10	10	100.00
P	3	3	100.00	14	14	100.00	19	19	100.00	20	19	95.00	20	20	100.00	19	19	100.00
30-39																		
F	1	1	100.00	8	8	100.00	10	10	100.00	9	9	100.00	9	8	88.89	10	10	100.00
M	9	9	100.00	9	8	88.89	10	9	90.00	10	10	100.00	10	10	100.00	10	9	90.00
P	10	10	100.00	17	16	94.12	20	19	95.00	19	19	100.00	19	18	94.74	20	19	95.00
40-49																		
F	1	1	100.00	10	10	100.00	10	10	100.00	10	9	90.00	8	8	100.00	10	10	100.00
M	9	9	100.00	10	9	90.00	10	9	90.00	10	10	100.00	10	9	90.00	10	10	100.00
P	10	10	100.00	20	19	95.00	20	19	95.00	20	19	95.00	18	17	94.44	20	20	100.00
50-59																		
F	0			8	8	100.00	10	9	90.00	10	9	90.00	9	8	88.89	10	10	100.00
M	11	10	90.91	9	9	100.00	10	10	100.00	10	10	100.00	10	10	100.00	10	10	100.00
P				17	17	100.00	20	19	95.00	20	19	95.00	19	18	94.74	20	20	100.00
60-69																		
F	4	4	100.00	10	9	90.00	10	10	100.00	10	8	80.00	8	8	100.00	10	10	100.00
M	9	8	88.89	8	8	100.00	11	11	100.00	10	9	90.00	10	10	100.00	10	10	100.00
P	13	12	92.31	18	17	94.44	21	21	100.00	20	17	85.00	18	18	100.00	20	20	100.00
70-79																		
F	3	3	100.00	9	9	100.00	10	10	100.00	10	10	100.00	10	8	80.00	9	9	100.00
M	12	12	100.00	10	9	90.00	10	10	100.00	10	10	100.00	10	10	100.00	9	9	100.00
P	15	15	100.00	19	18	94.74	20	20	100.00	20	20	100.00	20	18	90.00	18	18	100.00
80-89																		
F	1	1	100.00	10	10	100.00	10	10	100.00	10	10	100.00	10	10	100.00	10	9	90.00
M	9	9	100.00	3	2	66.67	6	6	100.00	10	9	90.00	10	10	100.00	10	9	90.00

P	10	10	100.00	13	12	92.31	16	16	100.00	20	19	95.00	20	20	100.00	20	18	90.00
90-99																		
F	2	1	50.00	0			3	3	100.00	7	7	100.00	6	6	100.00	2	2	100.00
M	0			1	1	100.00	1	1	100.00	0			9	8	88.89	4	4	100.00
P							4	4	100.00				15	14	93.33	6	6	100.00
100+																		
F	0			0			0			0			1	1	100.00	0		
M	0			0			0			0			3	3	100.00	0		
P	0						0						4	4	100.00			
Total	74	71	95.95	119	114	95.80	140	137	97.86	146	139	95.21	153	147	96.08	143	140	97.90

Table A3.3. Pelvis and skull combined sex assessment; numbers and percentages correct by age group and sex for each collection

F = female; M = male; P = sexes pooled; *n* = number of individuals; *n* corr = number of correct sex assignments; % corr = percentage correct sex assignments.

Appendix 4: Sex Differences in Ageing by Method – Within Collection

The equality of means and variances for each sex, divided by collection and score or phase, were tested in order to look for any sex differences. Independent sample t-tests and Levene's Test for Equality of Variances were used. P-values are reported here; those in slightly larger and bold type are significant. Values have been rounded to two decimal places; some values reported as 0.05 were slightly larger than 0.05 to three decimal places, and so have not been considered significant, while others reported as 0.05 were lower than 0.05 to three decimal places, and so have been considered significant.

	Phase I		Phase II		Phase III		Phase IV		Phase V		Phase VI	
	Var	μ	Var	μ	Var	μ	Var	μ	Var	μ	Var	μ
Coimbra	0.46	0.22	--	**0.03**	0.06	0.16	0.46	0.55	0.51	0.49	0.23	0.24
Dart	0.64	0.53	0.90	**0.02**	**0.02**	0.23	0.51	0.82	0.95	0.46	0.63	0.83
Grant	--	--	--	0.33	--	--	0.94	0.80	0.35	0.79	0.06	0.30
Lisbon	--	0.51	0.07	0.64	0.45	0.61	0.11	0.87	0.65	0.50	0.42	0.38
Pretoria	--	1.0	0.09	0.75	0.75	0.50	0.54	0.39	0.60	0.47	0.99	0.70
Spitalfields	--	0.87	--	0.33	0.19	0.97	0.77	0.30	0.11	0.96	0.18	0.74

Table A4.1. Suchey-Brooks pubic symphysis method: equality of means between males and females within collections
Var: equality of variance; μ: equality of means (t-test); bold type: the means or variances are not equal, there is a significant difference between them); --: not applicable (e.g. maybe only one case (or none) for one of the sexes, so Levene's test could not be calculated).

For the Suchey-Brooks pubic symphysis method, in only three instances were there statistically significant differences: the mean ages for phase II between males and females from the Dart Collection and from the Coimbra Collection, and in variance for phase III Dart Collection males and females (see Table A4.1). In the Coimbra Collection, only one male was categorised as phase II (so variance could not be calculated) – while this may seem to be an issue with age distribution, perhaps leading to the suspicion that higher numbers of younger individuals should be included, this is not the case, as sampling was (approximately) equal across age groups and collections. That is, ten individuals per age group for each sex were sampled for each collection, except where there were not enough individuals in a given age category – and the gaps are largely in the older age groups. Overall, the majority of individuals presented morphology placing them in phases IV and V.

The same was done for the Buckberry-Chamberlain auricular surface method (see Table A4.2); phase I was excluded as only three individuals were in this category – one Lisbon male, one Lisbon female (aged 20 and 21, respectively), and one Pretoria female (aged 22). Phases IV, V, and VI have the highest numbers individuals overall. Phase II showed significant differences in variance between Coimbra males and females and Pretoria males and females; Coimbra males and females displayed differences in mean nearing significance for phase III. Variance between Coimbra males and females and Dart males and females was significant for phase VII; no other significant sex differences were found using the Buckberry-Chamberlain method.

	Phase II		Phase III		Phase IV		Phase V		Phase VI		Phase VII	
	Var	μ	Var	μ	Var	μ	Var	μ	Var	μ	Var	μ
Coimbra	**0.01**	0.35	0.96	**0.05**	0.54	0.79	0.58	0.72	0.70	0.67	**0.04**	0.07
Dart	0.26	0.31	0.82	0.98	0.55	0.18	0.97	0.66	0.40	0.79	**0.04**	0.16
Grant	--	--	--	--	--	--	0.11	0.91	0.08	0.33	0.57	0.89
Lisbon	0.49	0.62	0.76	0.40	0.42	0.65	0.22	0.13	0.17	0.12	0.83	0.30
Pretoria	**0.04**	0.64	0.47	0.59	0.75	0.58	0.70	0.73	0.70	0.93	0.74	0.43
Spitalfields	0.66	0.61	0.49	0.68	0.52	0.88	0.73	0.79	0.75	0.79	0.45	0.76

Table A4.2. Buckberry-Chamberlain auricular surface method: equality of means between males and females within collections
Var: equality of variance; μ: equality of means (t-test); bold type: the means or variances are not equal, there is a significant difference between them); --: not applicable (e.g. maybe only one case (or none) for one of the sexes, so Levene's test could not be calculated).

The only significant differences found between the sexes using the Meindl-Lovejoy method (see Table A4.3) did not match those found using the Buckberry-Chamberlain method. Only Pretoria males and females in the 40-44 and 45-49 year phase categories showed significant differences in mean age; no other significant differences were found, though Dart males and females showed differences in mean in the 50-60 group nearing significance.

For lateral-anterior sutures scores, scores of 1 and 2 were not included in analysis for equality of variance and mean as there were not enough individuals in those categories – no individuals scored 1, and only one female from Coimbra, aged 26, had a score of 2. The only significant mean age sex differences found in terms of lateral-anterior sutures (see Table A4.4 for details) were between Lisbon males and females placed in Phase 5. Significant differences in variance between the sexes were found for Lisbon in phases 5 and 6. Differences nearing significance were found for phase 6 between Coimbra males and females, where variance and mean p-values were only marginally larger than 0.05 (.059 and 0.051, respectively). Variance and means between the sexes for phase 8 could not be calculated, as the few individuals belonging to phase 8 (from Grant, Spitalfields, Lisbon and Coimbra only) were all male.

	20-24		25-29		30-34		35-39		40-44		45-49		50-60		60+	
	Var	µ	Var	µ	Var	µ	Var	µ	Var	µ	Var	µ	Var	µ	Var	µ
Coimbra	--	--	.84	.36	.34	.77	.46	.81	.79	.49	.27	.94	.49	.98	--	--
Dart	.75	.83	.66	.83	.83	.70	.12	.20	.75	.45	.31	.66	.98	.06	--	.91
Grant	--	--	--	--	--	--	--	--	--	.24	.97	.89	.22	.78	.98	.12
Lisbon	.52	.69	.10	.46	.13	.58	.18	.66	.44	.10	.78	.11	.47	.17	--	--
Pretoria	.28	.72	.34	.92	.48	.54	.35	.92	.74	**.03**	.22	**.04**	.58	.88	--	--
Spitalfields	--	--	.74	.57	.85	.94	.16	.55	.52	.71	.63	.75	.18	.70	.35	.21

Table A4.3. Meindl-Lovejoy auricular surface method: equality of means between males and females within collections

Var: equality of variance; µ: equality of means (t-test); bold type: the means or variances are not equal, there is a significant difference between them); --: not applicable (e.g. maybe only one case (or none) for one of the sexes, so Levene’s test could not be calculated).

	3		4		5		6		7		8	
	Var	µ	Var	µ	Var	µ	Var	µ	Var	µ	Var	µ
Coimbra	--	0.83	--	.74	.95	.61	.06	.05	.57	.76	--	--
Dart	--	--	--	--	.24	.83	.27	.19	.58	.52	--	--
Grant	--	--	--	--	--	--	.15	.23	.38	.20	--	--
Lisbon	--	--	.27	.18	**.04**	**.02**	**.05**	.10	.37	.39	--	--
Pretoria	--	--	--	--	.12	.79	.32	.44	.67	.82	--	--
Spitalfields	--	--	--	--	--	--	.17	.36	.85	.95	--	--

Table A4.4. Lateral-anterior suture phases: equality of means between males and females within collections
Var: equality of variance; µ: equality of means (t-test); bold type: the means or variances are not equal, there is a significant difference between them); --: not applicable (e.g. maybe only one case (or none) for one of the sexes, so Levene's test could not be calculated).

The vault suture phases also showed few significant sex differences (see Table A4.5). A significant difference in mean age between Coimbra males and females in phase 3 was found, and between Dart males and females in phase 6. Variance between Lisbon males and females in phase 3 is also significantly different, while the difference in means for phase 6 for Lisbon males and females were nearing significance, with a p-value of 0.053.

	2		3		4		5		6		7	
	Var	µ	Var	µ	Var	µ	Var	µ	Var	µ	Var	µ
Coimbra	--	--	.07	**.03**	.47	.13	.11	.20	.96	.29	.77	.94
Dart	.73	.12	.09	.09	.57	.09	.33	.84	.29	**.01**	.90	.54
Grant	--	--	--	--	--	--	--	.75	--	--	--	--
Lisbon	--	--	**.02**	.21	.90	.31	.76	.45	.51	.05	.50	.42
Pretoria	--	--	.09	.52	.21	.39	.24	.88	.17	.59	.15	.67
Spitalfields	--	--	--	.56	.94	.49	.34	.88	.36	.14	.70	.41

Table A4.5. Vault suture scores: equality of means between males and females within collections
Var: equality of variance; µ: equality of means (t-test); bold type: the means or variances are not equal, there is a significant difference between them); --: not applicable (e.g. maybe only one case (or none) for one of the sexes, so Levene's test could not be calculated).

Calculations for equality of variance and mean were not done for phases 1 to 4, and 13 to 15 of İşcan and Loth's fourth rib method, as there were either no individuals with those scores, or only individuals of one sex. Some significant differences were found; Coimbra males and females showed significant differences in variance for phase 9, and in mean for phase 10, as did Lisbon males and females, and Pretoria males and females showed significant differences in variance for phases 6 and 10 and in mean for phase 8 (see Table A4.6 for full details).

	5		6		7		8		9		10		11		12	
	Var	µ	Var	µ	Var	µ	Var	µ	Var	µ	Var	µ	Var	µ	Var	µ
Coimbra	--	--	--	--	--	.74	.11	.82	**.01**	.44	.13	**.00**	--	.43	--	--
Dart	--	1.0	--	.51	.64	.63	.79	.88	--	--	--	.89	.73	.78	--	.57
Grant	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Lisbon	--	--	--	--	.09	.30	.48	.31	.72	.89	.37	**.02**	--	.82	--	--
Pretoria	--	--	**.03**	.25	.22	.65	.42	**.05**	.24	.98	**.04**	.12	--	.48	--	--
Spitalfields	--	--	.12	.36	.15	.31	.37	.61	.82	.17	.36	.45	.95	.24	--	--

Table A4.6. İşcan-Loth rib phases: equality of means between males and females within collections

Var: equality of variance; µ: equality of means (t-test); bold type: the means or variances are not equal, there is a significant difference between them); --: not applicable (e.g. maybe only one case (or none) for one of the sexes, so Levene's test could not be calculated).

A.4.1 Buckberry-Chamberlain Scored Traits

The scored traits for the Buckberry-Chamberlain method were each subjected to tests of equality of variance and mean by sex for each collection as well. As with the phases for each method, no clear trends were found in terms of significant differences between the sexes for any particular collection or scored trait.

For transverse organisation, only Spitalfields had enough males and females with scores of 1 to calculate variance and mean. Meanwhile, for the Grant Collection, most individuals scored either 4 or 5; only the mean for Grant for scores of 2 could be calculated, while no calculations were possible for scores of 3 or 1 (as not enough males and females had these scores). The only significant differences were in mean age for scores of 5 between Coimbra males and females and Spitalfields males and females (see Table A4.7 for details). For surface texture scores, the only significant differences were in variance between Coimbra males and females with scores of 1, and Dart males and females with scores of 5 (see Table A4.8). No mean age differences were significant. For microporosity, there are only three possible scores; out of these, only Dart males and females with scores of 2 showed a significant difference in mean age; no other significant sex differences in microporosity scores were found (see Table A4.9 for details). For macroporosity, again, with three possible scores, no sex differences were found in mean age or variance (see Table A4.10). The last morphological trait for the Buckberry-Chamberlain method, apical change, did show some significant sex differences (see Table A4.11). Significant differences in variance and mean were found between Dart males and females with scores of 1, as well as significant differences in variance between Pretoria males and females with scores of 1, and mean between Coimbra males and females with scores of 2. Mean age differences between Pretoria males and females with scores of 1 and 2 also approach significance.

	1		2		3		4		5	
	Var	μ	Var	μ	Var	μ	Var	μ	Var	μ
Coimbra	--	--	.10	.18	.93	.60	.86	.82	.29	**.02**
Dart	--	--	.99	.97	.46	.67	.45	.91	.57	.44
Grant	--	--	--	.29	--	--	.62	.38	.44	.80
Lisbon	--	--	.27	.46	.57	.54	.33	.06	.44	.74
Pretoria	--	--	.10	.86	.06	.08	.61	.53	.91	.48
Spitalfields	.30	.84	.22	.37	.85	.83	.95	.44	.14	**.00**

Table A4.7. Buckberry-Chamberlain auricular surface method: transverse organisation score, equality of means between males and females within collections
Var: equality of variance; μ: equality of means (t-test); bold type: the means or variances are not equal, there is a significant difference between them); --: not applicable (e.g. maybe only one case (or none) for one of the sexes, so Levene's test could not be calculated).

	1		2		3		4		5	
	Var	μ	Var	μ	Var	μ	Var	μ	Var	μ
Coimbra	**.04**	.34	.61	.49	.68	.38	.71	.23	--	--
Dart	--	.24	.72	.96	.23	.35	.49	.78	**.02**	.34
Grant	--	--	--	.20	--	--	.95	.35	.76	.66
Lisbon	.45	.92	.09	.65	.42	.51	.79	.19	--	--
Pretoria	--	--	.15	.24	.15	.84	.79	.34	--	--
Spitalfields	--	--	.94	.60	.08	.12	.10	.33	.34	.96

Table A4.8. Buckberry-Chamberlain auricular surface method: surface texture score, equality of means between males and females within collections

Var: equality of variance; μ: equality of means (t-test); bold type: the means or variances are not equal, there is a significant difference between them); --: not applicable (e.g. maybe only one case (or none) for one of the sexes, so Levene's test could not be calculated).

	1		2		3	
	Var	μ	Var	μ	Var	μ
Coimbra	--	.10	.32	.97	.38	.33
Dart	.40	.96	.37	**.02**	.91	.38
Grant	--	.58	.13	.24	.82	.44
Lisbon	.31	.62	.77	.61	.30	.48
Pretoria	.42	.13	.50	.64	.34	.57
Spitalfields	--	--	.93	.76	.41	.59

Table A4.9. Buckberry-Chamberlain auricular surface method: microporosity score, equality of means between males and females within collections

Var: equality of variance; μ: equality of means (t-test); bold type: the means or variances are not equal, there is a significant difference between them); --: not applicable (e.g. maybe only one case (or none) for one of the sexes, so Levene's test could not be calculated).

	1		2		3	
	Var	μ	Var	μ	Var	μ
Coimbra	.18	.66	.52	.86	.42	.21
Dart	.52	.25	.88	.23	.97	.29
Grant	.93	.32	.28	.85	.80	.33
Lisbon	.57	.56	.86	.91	.93	.20
Pretoria	.72	.60	.72	.92	.31	.84
Spitalfields	.73	.35	.10	.85	.32	.40

Table A4.10. Buckberry-Chamberlain auricular surface method: macroporosity score, equality of means between males and females within collections

Var: equality of variance; μ: equality of means (t-test); bold type: the means or variances are not equal, there is a significant difference between them); --: not applicable (e.g. maybe only one case (or none) for one of the sexes, so Levene's test could not be calculated).

	1		2		3	
	Var	μ	Var	μ	Var	μ
Coimbra	.56	.94	.93	**.04**	.28	.64
Dart	**.01**	**.01**	.78	.57	.81	.85
Grant	--	.19	.82	.11	.54	.66
Lisbon	.70	.92	.86	.21	.14	.52
Pretoria	**.05**	.05	.17	.05	.53	.26
Spitalfields	.18	.22	.34	.79	.82	.42

Table A4.11. Buckberry-Chamberlain auricular surface method: apical changes score, equality of means between males and females within collections

Var: equality of variance; μ: equality of means (t-test); bold type: the means or variances are not equal, there is a significant difference between them); --: not applicable (e.g. maybe only one case (or none) for one of the sexes, so Levene's test could not be calculated).

Appendix 5: Variation in Phase/Score Distribution

A.5.1 Variation in Phase Distribution

A.5.1.1 Suchey-Brooks Pubic Symphysis Method

Significant differences in median for the Suchey-Brooks pubic symphysis phase distribution were found for females only (Table A5.1) between Grant and every other collection. However, Grant's female age distribution was skewed towards the older end of the age range, as few females in general were available for study from the Grant Collection, and the majority were over 60 years old. Accordingly, the Grant female phase distribution is skewed towards the higher phases.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.999	.471	.268	.037	.808	.474	1.000	.525	.350	.361
Dart			.098	.012	.951	.966	1.000	.942	.657	.721
Grant					.239	.024	.123	.014	.060	.020
Lisbon							.951	.910	.731	.970
Pretoria									.657	.681

Table A5.1. Suchey-Brooks, females only: two-sample K-S and MWU results

A.5.1.2 Vault Suture Closure Method

For females, the only significant differences were in median between Dart compared to Coimbra and Spitalfields. Results for the statistical tests are below, in Table A5.2.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.129	.049	--	.865	.973	.732	.746	.758	.925	.465
Dart			--	.583	.263	.062	.497	.051	.065	.007
Grant					--	.778	--	.750	--	.898
Lisbon							1.000	1.000	.708	.285
Pretoria									.523	.276

Table A5.2. Vault sutures, females only: two-sample K-S and MWU results

A.5.1.3 Sternal End of Fourth Rib Method

No significant differences found between collections for the sternal end of the fourth rib; the distributions between Dart and Lisbon neared significance. Reasons for this difference were discussed in Chapter 4. Tables A5.3 to A5.5 provide all p-values for the K-S and MWU tests. Table A5.6a and A5.6b (split into two) provides phase frequencies by age group, for females, males and the sexes pooled.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.797	.645	.824	.751	.553	.342	1.000	.991	.511	.126
Dart			.833	.843	.037	.114	.499	.593	.109	.059
Grant					.791	.461	.835	.696	.563	.410
Lisbon							.741	.375	.916	.472
Pretoria									.732	.109

Table A5.3. Fourth rib, sexes pooled: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.994	.697	--	--	1.000	.711	.999	.598	.999	.904
Dart			--	--	.941	.534	.816	.497	.997	.745
Grant					--	--	--	--	--	--
Lisbon							1.000	.917	.998	.787
Pretoria									1.000	.746

Table A5.4. Fourth rib, females only: two-sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.997	.979	.734	.554	.591	.280	1.000	.604	.896	.434
Dart			.559	.357	.212	.184	.804	.572	.685	.477
Grant					.998	.883	.874	.820	.919	1.000
Lisbon							.914	.686	.933	.991
Pretoria									1.000	.754

Table A5.5. Fourth rib, males only: two-sample K-S and MWU results

	3			4			5			6			7			8		
	F	M	P	F	M	P	F	M	P	F	M	P	F	M	P	F	M	P
Coimbra	0	1	**1**	0	1	**1**	3	0	**3**	4	0	**4**	2	2	**4**	7	4	**11**
Dart	1	0	**1**	1	0	**1**	2	2	**4**	2	1	**3**	3	2	**5**	3	2	**5**
Grant	0	0	**0**	0	0	**0**	0	0	**0**	0	0	**0**	**0**	0	**0**	0	0	**0**
Lisbon	0	1	**1**	0	0	**0**	2	0	**2**	4	0	**4**	6	2	**8**	5	3	**8**
Pretoria	0	0	**0**	0	0	**0**	1	1	**2**	6	3	**9**	3	4	**7**	5	6	**11**
Spitalfields	1	0	**1**	2	0	**2**	2	0	**2**	8	2	**10**	7	2	**9**	5	5	**10**

Table A5.6a. Fourth rib phase frequency by collection for each sex and the sexes pooled: phases 3 to 8
F: female; M: male; P: pooled.

	9			10			11			12			13			14			15		
	F	M	P	F	M	P	F	M	P	F	M	P	F	M	P	F	M	P	F	M	P
Coimbra	2	4	**6**	2	4	**6**	1	8	**9**	0	4	**4**	2	0	**2**	0	2	**2**	0	0	**0**
Dart	0	3	**3**	1	4	**5**	2	14	**16**	1	3	**4**	0	2	**2**	0	0	**0**	0	0	**0**
Grant	0	1	**1**	0	1	**1**	0	0	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lisbon	4	5	**9**	3	6	**9**	1	4	**5**	0	2	**2**	0	1	**1**	0	0	**0**	0	0	**0**
Pretoria	4	5	**9**	3	5	**8**	1	9	**10**	0	2	**2**	0	1	**1**	0	4	**4**	0	2	**2**
Spitalfields	4	2	**6**	5	2	**7**	4	4	**8**	0	3	**3**	0	2	**2**	0	0	**0**	0	0	**0**

Table A5.6b. Fourth rib phase frequency by collection for each sex and the sexes pooled: phases 9 to 15
F: female; M: male; P: pooled.

A.5.2 Buckberry-Chamberlain Scored Traits

A.5.2.1 Macroporosity

Compared to the macroporosity results for the sexes pooled, relatively few significant differences were found when the sexes were separated. For females only, significant differences in score distribution and median were found in Coimbra compared to Lisbon and Pretoria, and in median only in Spitalfields compared to Lisbon and Pretoria. For males only, significant differences were found in median only between Coimbra and Dart, and in Grant compared to Dart and Pretoria. The differences in score frequency causing these differences are described in Chapter 4.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.105	.187	.353	.444	.017	.015	.029	.015	1.00	.991
Dart			.725	.211	.989	.372	.923	.375	.339	.229
Grant					.316	.073	.248	.070	.620	.477
Lisbon							1.00	.990	.082	.027
Pretoria									.123	.026

Table A5.7. Buckberry-Chamberlain auricular surface, females only, macroporosity score: two - sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.106	.026	.739	.518	.385	.133	.146	.065	1.00	.617
Dart			.158	.010	1.00	.523	1.00	.801	.352	.104
Grant					.498	.055	.209	.027	.550	.296
Lisbon							1.00	.720	.785	.340
Pretoria									.428	.197

Table A5.8. Buckberry-Chamberlain auricular surface, males only, macroporosity score: two - sample K-S and MWU results

A.5.3 Within-Collection Distribution Variation by Sex

Table A5.9 presents the p-values for differences between males and females for each collection and for each age determination method. Where the data were too few, certain collections could not be tested – for example, the Grant Collection had rib data for only two males.

The Suchey-Brooks method, which does have separate standards for males and females, was tested first. Only the Coimbra Collection had significant differences in both phase distribution and median between males and females. This was due to lower proportions of females in phases III and V compared to males, but higher proportions of females in phases IV and VI; the higher proportion of females in phase VI probably also contributed to their higher median. A significant difference in Suchey-Brooks median was found between the sexes for the Grant Collection, as more males were in phases IV, V, and VI; however, the small female sample size also affected the median, so the difference in median here should not be considered a biologically meaningful result. No other significant differences were found between the sexes of other collections for the Suchey-Brooks phase distributions and medians.

A significant difference between the sexes in median in the Grant Collection was found for the Buckberry-Chamberlain method; however, the small female sample size and few females of younger ages were likely the determining factors in the higher female median, rather than any meaningful biological characteristics. Similarly, the only significant differences between sexes when the Meindl-Lovejoy auricular surface data were considered were in the Grant Collection, in both phase distribution and median. However, the reasons for the difference in median were the same as described above for the Buckberry-Chamberlain differences; the distribution difference was likely also affected by the relatively high proportion of Grant females in the 60+ phase and lower proportions of females in the lower phases, but was at least partially as a result of the few younger females in the sample.

The lateral-anterior cranial sutures showed significant differences in phase distribution and median between Lisbon males and females. Despite the fact that there were seven females in the 90 to 99 age group and no males, Lisbon females had higher proportions in phases 3 to 6, while males had higher proportions in phases 7 and 8, resulting in the distribution differences and higher male median. Significant differences in median were also found between Pretoria and Spitalfields males and females. The Pretoria individuals all had lateral-anterior suture scores that were confined to phases 5, 6, and 7. The males, however, had a higher median due to a higher proportion in phase 7 compared to the females. Similarly, for Spitalfields, there were higher proportions of females in phases 5 and 6, and a higher proportion of males in phases 7 and 8, leading to a higher male median. For vault sutures, the significant differences were also in median only, between Dart and Spitalfields males and females. For Dart, the higher male median was a result of higher proportions of males in phases 5, 6 and 7 compared to the females. Similarly, for Spitalfields, there were higher female proportions in phases 3 and 4, and higher male proportions in phases 5 and 6, resulting in a higher male median.

There were sex differences in all collections in the fourth rib (except Grant, as there were too few individuals to test), in both phase distribution and median for Coimbra, Dart, Lisbon and Pretoria, and in median only for Spitalfields. The Spitalfields difference in median was the result of a lower female median, as more females were in the lower phases (3 to 7), with higher proportions of males in the higher phases. The differences in Lisbon males and females were for the same reasons (but to a higher degree, resulting in distributional differences as well as median); higher proportions of females were in the lower phases, from phase 4 to 8, after which the proportions change, and male proportions were higher from phase 9 onwards. The same was true for Coimbra and Dart males and females. A similar situation occurs with Pretoria males and females, but the "switch" seemed to occur later, around phase 10; however, the female sample size for Pretoria was smaller than that of the male (nearly half the size), so the precise point is more difficult to determine.

Within-collection sex differences in score distributions were also analysed for each of the Buckberry-Chamberlain auricular surface skeletal traits that are scored individually. Few sex differences were found in these traits (see Table A5.10, below). For apical change, a significant difference in both score distribution and median were found between Dart males and females. There were a higher proportion of Dart females with apical change scores of 1, and higher proportions of Dart males with scores of 2 and 3, leading to the distribution differences and higher male median. Pretoria males and female also presented a significant difference, but in median only; again, the male median was higher due to their higher proportions of scores of 3 (and 2, to a lesser extent), and a higher female proportion of scores of 1. For both surface texture and transverse organisation scores, the only significant differences were in median for Grant males and females. While for both surface texture and transverse organisation, Grant females had much higher proportions of high scores (4 and 5), the small female sample size and few young females affected the median value; thus, these significant differences must be viewed with caution, as with the sex differences in ageing method median described earlier. The last significant difference found in any of the Buckberry-Chamberlain auricular surface features was the median in the macroporosity scores of Dart males and females. The Dart difference in median was a result of a lower proportion of females with scores of 1 (and to a lesser extent, scores of 2), and a higher proportion of females with scores of 3.

	S-B		B-C		M-L		L-A		V		R	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	**.045**	**.002**	1.00	.829	.626	.959	.949	.237	.400	.146	**.003**	**.000**
Dart	.773	.830	.508	.796	.840	.440	.582	.156	.059	**.006**	**.004**	**.002**
Grant	.126	**.046**	.097	**.025**	**.024**	**.005**	.157	.137	--	--	--	--
Lisbon	.423	.513	.581	.429	.895	.710	**.006**	**.000**	.733	.110	**.022**	**.001**
Pretoria	1.00	.746	.998	.510	.670	.065	.190	**.025**	1.00	.752	**.024**	**.001**
Spitalfields	.963	.951	1.00	.611	.560	.711	.209	**.006**	.068	**.032**	.073	**.002**

Table A5.9. Age determination methods, within-collection phase distribution, males compared to females: two -sample K-S and MWU results
S-B: Suchey-Brooks pubic symphysis method; B-C: Buckberry-Chamberlain auricular surface method; M-L: Meindl-Lovejoy auricular surface method; L-A: lateral-anterior suture method; V: vault suture method; R: fourth rib method.

	AC		ST		TO		MI		MA	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.771	.200	1.000	.885	1.000	.396	1.000	.479	.673	.090
Dart	**.017**	**.001**	1.000	.776	.865	.160	.937	.357	.278	**.037**
Grant	1.000	.519	.206	**.018**	.247	**.050**	1.000	.936	.643	.299
Lisbon	.889	.822	1.000	.967	1.000	.768	.917	.193	1.000	.583
Pretoria	.162	**.004**	1.000	.863	.413	.069	.700	.191	.986	.363
Spitalfields	1.000	.693	.587	.628	1.000	.594	1.000	.550	.547	.056

Table A5.10. Buckberry-Chamberlain auricular surface features, within-collection score distribution, males compared to females: two -sample K-S and MWU results
AC: apical changes; ST: surface texture; TO: transverse organisation; MI: microporosity; MA: macroporosity.

Appendix 6: Variation in Phase/Score by Age Group

The Kolomogorov-Smirnov test and Mann-Whitney U test were used to analyse variation in distribution and median, for the sexes pooled (group sizes for either sex alone would be quite small).

A.6.1 Suchey-Brooks Phase Distribution by Age Group (K-S and MWU Tests)

The Suchey-Brooks phase distributions were analysed first for each age group (see Tables A6.1a to A6.1h). Some significant differences in phase distribution and median were found in various age groups, but no consistent differences between particular collections were found. For instance, differences in distribution and phase nearing significance were found between Dart and Lisbon in the 40 to 49 age group, but the only other near-significant difference between these two collections was in median only in the 70 to 79 age group. Another example was the significant differences between Spitalfields and every other collection except Lisbon for the 40 to 49 age group – Spitalfields had no other significant differences in any other age group. This variability seems to suggest that stochastic variation is at work rather than biological differences; however, perhaps research that involved fewer collections but larger sample sizes from each collection would be able to provide more information.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.372	.067	.834	.277	.985	.591	.794	.159	1.000	.939
Dart			.997	1.000	.375	.104	.991	.741	.744	.171
Grant					.862	.391	1.000	.867	.918	.337
Lisbon							.826	.279	.997	.744
Pretoria									.992	.289

Table A6.1a. Suchey-Brooks, sexes pooled, age 20-29: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.218	.022	.873	.718	.559	.065	.993	.348	.873	.310
Dart			.383	.064	.443	.482	.553	.112	.478	.150
Grant					.361	.101	.843	.587	.946	.460
Lisbon							.997	.329	.989	.349
Pretoria									1.000	.926

Table A6.1b. Suchey-Brooks, sexes pooled, age 30-39: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.300	.076	.288	.059	.972	.378	.655	.242	.526	.044
Dart			.956	.391	.028	.013	1.000	.549	.028	**.002**
Grant					.043	.021	.946	.225	.043	.006
Lisbon							.094	.059	.526	.132
Pretoria									.094	.005

Table A6.1c. Suchey-Brooks, sexes pooled, age 40-49: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.906	.319	.032	1.000	.488	1.000	.570	1.000	.847
Dart			.383	.036	1.000	.419	1.000	.491	1.000	.774
Grant					.134	.014	.134	.014	.225	.062
Lisbon							1.000	.879	1.000	.678
Pretoria									1.000	.776

Table A6.1d. Suchey-Brooks, sexes pooled, age 50-59: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.559	.088	1.000	1.000	.747	.272	1.000	.955	.965	.583
Dart			.713	.124	1.000	.547	.804	.137	.999	.307
Grant					.867	.322	1.000	.957	.992	.622
Lisbon							.939	.367	1.000	.657
Pretoria									1.000	.681

Table A6.1e. Suchey-Brooks, sexes pooled, age 60-69: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.560	.029	1.000	.797	1.000	.834	.978	.741	.939	.879
Dart			.202	.042	.890	.025	.329	.164	.097	.144
Grant					.954	.690	1.000	.610	1.000	.916
Lisbon							.890	.808	.714	.808
Pretoria									.999	.781

Table A6.1f. Suchey-Brooks, sexes pooled, age 70-79: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.895	1.000	.593	.999	.448	1.000	.377	.870	.473
Dart			1.000	.596	.841	.424	.983	.355	.521	.349
Grant					1.000	.871	1.000	.811	.594	.298
Lisbon							1.000	.923	.521	.214
Pretoria									.337	.161

Table A6.1g. Suchey-Brooks, sexes pooled, age 80-89: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.651	--	.414	1.000	.807	1.000	1.000	--	.414
Dart			--	.273	.921	.463	.999	.639	--	.273
Grant					--	.533	--	.450	--	1.000
Lisbon							1.000	.805	--	.533
Pretoria									--	.450

Table A6.1h. Suchey-Brooks, sexes pooled, age 90-99: two -sample K-S and MWU results

A.6.2 Buckberry-Chamberlain Phase Distribution by Age Group (K-S and MWU Tests)

The Buckberry-Chamberlain method presented more significant differences by age group (Tables A6.2a to A6.2h). While significant differences in phase distribution and median appeared throughout the age ranges, few were consistent. Significant or near-significant differences in median and, for some age groups, distribution as well, did occur consistently between Grant and Pretoria. These began at the 20 to 29 age group, remaining through to the 80 to 89 age group. For all age groups, Grant's median was higher and its distribution was more focused on higher phases compared to Pretoria. Grant's significant differences compared to other collections, where they appeared throughout the age groups, were for the same reasons. Where Spitalfields showed significant differences compared to the other collections, its median was also higher and distribution focused on higher phases, but was less extreme than Grant. The South African and Portuguese collections had lower medians and phase distributions more focused on lower phases.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.709	.211	.322	.021	.256	.030	1.000	.503	.437	.175
Dart			.375	.080	.059	**.002**	.560	.073	1.000	.790
Grant					.065	**.003**	.266	.014	.429	.132
Lisbon							.735	.148	.085	.005
Pretoria									.327	.073

Table A6.2a. Buckberry-Chamberlain, sexes pooled, age 20-29: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.142	.144	.240	.109	.227	.009	.132	.016	.968	.941
Dart			.222	.087	.526	.988	.446	.876	.794	.274
Grant					.007	.005	.029	.011	.471	.180
Lisbon							1.000	.951	.300	.055
Pretoria									.184	.084

Table A6.2b. Buckberry-Chamberlain, sexes pooled, age 30-39: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.654	.388	.063	.896	.534	.807	.532	.735	.079
Dart			.236	.093	.659	.304	.539	.245	.945	.155
Grant					.288	.019	.215	.009	.555	.496
Lisbon							.985	.738	.084	.026
Pretoria									.050	.006

Table A6.2c. Buckberry-Chamberlain, sexes pooled, age 40-49: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.172	.010	.979	.614	.978	.460	.055	**.003**	.911	.308
Dart			.291	.048	.329	.104	.990	.466	.025	.007
Grant					.988	.812	.127	.022	.519	.230
Lisbon							.449	.027	.571	.133
Pretoria									.048	**.002**

Table A6.2d. Buckberry-Chamberlain, sexes pooled, age 50-59: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.343	.015	.018	.006	.707	.196	.090	.006	1.000	.536
Dart			.012	**.000**	.840	.193	1.000	.853	.136	.008
Grant					.030	**.003**	.006	**.000**	.078	.026
Lisbon							.398	.102	.270	.104
Pretoria									.069	**.003**

Table A6.2e. Buckberry-Chamberlain, sexes pooled, age 60-69: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.256	.024	.996	.692	.941	.427	.211	.019	.557	.282
Dart			.285	.027	.645	.092	1.000	1.000	.022	**.002**
Grant					.497	.232	.297	.020	.990	.777
Lisbon							.560	.077	.497	.040
Pretoria									.032	**.001**

Table A6.2f. Buckberry-Chamberlain, sexes pooled, age 70-79: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.573	.083	.165	.043	.988	.348	.400	.121	.501	.050
Dart			**.003**	**.001**	.978	.507	1.000	.895	.015	**.000**
Grant					.032	.008	.015	**.003**	1.000	.636
Lisbon							.978	.514	.134	.005
Pretoria									.027	**.001**

Table A6.2g. Buckberry-Chamberlain, sexes pooled, age 80-89: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.723	.128	1.000	.480	.643	.144	.305	.047	--	.221
Dart			.465	.131	.983	.666	.983	.510	--	1.000
Grant					.405	.109	.203	.068	--	.157
Lisbon							.938	.241	--	.801
Pretoria									--	.640

Table A6.2h. Buckberry-Chamberlain, sexes pooled, age 90-99: two -sample K-S and MWU results

A.6.3 Meindl-Lovejoy Phase Distribution by Age Group (K-S and MWU Tests)

The Meindl-Lovejoy results by age group were examined next (see Tables A6.3a to A6.3h). While significant differences in distribution and median appear throughout the age range, the most interesting are centred around Grant and Spitalfields compared to the other collections. These began to appear in the 30 to 39 age group, and were most numerous in the 40 to 49 and 60 to 69 (Grant compared to most other collections) and 80 to 89 (Spitalfields compared to most other collections) age groups. The differences were because of Grant and Spitalfields' higher medians and distributions with more individuals in higher phases compared to the other collections; this was true of all the significant differences in all age groups in Grant and Spitalfields compared to the other collections.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.545	.066	.518	.213	.682	.161	.971	.853	.291	.029
Dart			.749	.529	.173	.005	.938	.128	1.000	.657
Grant					.502	.112	.506	.248	.759	.656
Lisbon							.708	.201	.089	.004
Pretoria									.814	.076

Table A6.3a. Meindl-Lovejoy, sexes pooled, age 20-29: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.567	.246	.056	.021	.303	.017	.941	.184	.957	.201
Dart			.058	.017	.412	.601	.981	.868	.104	.054
Grant					**.000**	**.000**	.006	**.001**	.367	.198
Lisbon							.994	.345	.056	**.001**
Pretoria									.349	.015

Table A6.3b. Meindl-Lovejoy, sexes pooled, age 30-39: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.093	.012	**.003**	**.000**	.917	.218	.722	.125	.013	**.000**
Dart			.043	.005	.793	.234	.854	.214	.513	.183
Grant					.005	**.000**	**.001**	**.000**	.124	.042
Lisbon							1.000	.913	.297	.017
Pretoria									.188	.008

Table A6.3c. Meindl-Lovejoy, sexes pooled, age 40-49: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.305	.034	.594	.307	.384	.947	.675	.089	.079	.041
Dart			.279	.012	.124	.163	1.000	.862	.019	**.001**
Grant					.329	.400	.594	.032	.820	.425
Lisbon							.343	.257	.325	.076
Pretoria									.076	**.002**

Table A6.3d. Meindl-Lovejoy, sexes pooled, age 50-59: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.283	.012	.015	**.000**	.997	.566	.222	.060	.643	.115
Dart			**.000**	**.000**	.172	.038	.441	.229	.017	**.001**
Grant					**.001**	**.000**	**.000**	**.000**	.353	.026
Lisbon							.775	.162	.157	.047
Pretoria									.015	**.002**

Table A6.3e. Meindl-Lovejoy, sexes pooled, age 60-69: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.020	**.003**	.869	.235	.329	.224	.381	.037	.987	.781
Dart			**.000**	**.000**	.127	.009	.709	.245	.004	**.001**
Grant					.226	.010	.035	**.002**	.999	.362
Lisbon							.731	.226	.484	.079
Pretoria									.142	.016

Table A6.3f. Meindl-Lovejoy, sexes pooled, age 70-79: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.396	.072	.474	.134	.978	.423	.081	.028	.051	.007
Dart			.016	**.002**	.897	.421	1.000	.555	**.000**	**.000**
Grant					.183	.041	**.002**	**.001**	.566	.202
Lisbon							.345	.222	.010	**.002**
Pretoria									**.000**	**.000**

Table A6.3g. Meindl-Lovejoy, sexes pooled, age 80-89: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.095	.117	.320	.094	.423	.049	.297	.039	--	.206
Dart			.230	.069	.591	.442	.983	.433	--	.633
Grant					.089	.035	.089	.031	--	.157
Lisbon							.203	.262	--	.817
Pretoria									--	.637

Table A6.3h. Meindl-Lovejoy, sexes pooled, age 90-99: two -sample K-S and MWU results

A.6.4 Lateral-Anterior Cranial Suture Phase Distribution by Age Group (K-S and MWU Tests)

Cranial sutures were examined next, starting with the lateral-anterior sutures. Some significant or near-significant differences were found, mostly in median only (see Tables A6.4a to A6.4h). The most interesting differences were in the youngest age groups. Spitalfields was found to have a significantly different (or nearly significantly different) median compared to Coimbra and Lisbon for the 20 to 29 age group, compared to all other collections except Grant for the 30 to 39 age group, and compared to all other collections except Coimbra for the 40 to 49 age group. In all cases, Spitalfields had the higher median and more highly peaked distribution compared to the other collections. While this provides some evidence supporting a faster rate of ageing for Spitalfields compared to the other collections, mean may be a better measure than median, and distributions were not significantly different. No other patterns were found for the lateral-anterior suture closure.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.333	.062	.994	.532	1.000	.976	.090	.006	.073	.009
Dart			1.000	.772	.548	.025	.654	.272	.432	.153
Grant					1.000	.406	.999	.350	.966	.216
Lisbon							.015	**.001**	.017	**.003**
Pretoria									1.000	.510

Table A6.4a. Lateral-anterior sutures, sexes pooled, age 20-29: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.883	1.000	.772	1.000	.855	.755	.442	.393	.030
Dart			1.000	.889	1.000	.746	.819	.581	.477	.044
Grant					1.000	.664	.952	.684	.689	.081
Lisbon							.484	.338	.246	.025
Pretoria									.314	.045

Table A6.4b. Lateral-anterior sutures, sexes pooled, age 30-39: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.463	.919	.208	.738	.212	1.000	.393	.507	.086
Dart			.995	.446	.953	.524	1.000	.883	.259	.021
Grant					1.000	.846	1.000	.517	.236	.020
Lisbon							1.000	.605	.203	.015
Pretoria									.259	.020

Table A6.4c. Lateral-anterior sutures, sexes pooled, age 40-49: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.567	1.000	.711	.999	.409	1.000	.614	1.000	.763
Dart			1.000	.915	1.000	.836	.994	.294	1.000	.818
Grant					1.000	.786	1.000	.436	1.000	.923
Lisbon							.892	.184	1.000	.676
Pretoria									1.000	.449

Table A6.4d. Lateral-anterior sutures, sexes pooled, age 50-59: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.115	.011	.512	.188	.982	.486	.152	.017	1.000	.769
Dart			1.000	.367	.778	.139	1.000	.764	.286	.033
Grant					.999	.628	1.000	.490	.771	.293
Lisbon							.869	.196	1.000	.645
Pretoria									.356	.046

Table A6.4e. Lateral-anterior sutures, sexes pooled, age 60-69: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.978	.500	.654	.402	.819	.472	.888	.349	.993	.346
Dart			1.000	.627	1.000	.826	1.000	.715	.872	.118
Grant					1.000	.791	1.000	.875	.500	.119
Lisbon							1.000	.894	.648	.125
Pretoria									.727	.080

Table A6.4f. Lateral-anterior sutures, sexes pooled, age 70-79: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.991	1.000	1.000	.881	1.000	.874	1.000	.812	.788	.159
Dart			.934	.896	1.000	.851	.963	.833	.917	.210
Grant					.998	1.000	1.000	.704	.664	.132
Lisbon							1.000	.705	.687	.183
Pretoria									.888	.206

Table A6.4g. Lateral-anterior sutures, sexes pooled, age 80-89: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.756	.059	.139	.046	.377	.048	.998	.221	--	.617
Dart			.528	.117	.997	.321	.578	.085	--	.456
Grant					.938	.513	.100	.008	--	.157
Lisbon							.242	.035	--	.339
Pretoria									--	1.000

Table A6.4h. Lateral-anterior sutures, sexes pooled, age 90-99: two -sample K-S and MWU results

A.6.5 Vault Cranial Suture Phase Distribution by Age Group (K-S and MWU Tests)

For the vault sutures, not enough Grant individuals with observable vault sutures were available in the 20 to 29, 30 to 39, 50 to 59, 70 to 79 or 90 to 99 age groups. Some significant differences in distribution and median were found, but these did not appear to follow any pattern (see Tables A6.5a to A6.5h). No differences were consistently found between any two collections.

The 40 to 49 age group had the most significant or near-significant differences, in phase distribution and median in Dart compared to Grant and Lisbon and between Grant and Pretoria. Further differences in this age group in median only were found in Grant compared to Lisbon and Spitalfields. Grant differs from the other collections due to an extremely limited range of phases for this age group (individuals are only in phases 6 and 7), while individuals from other collections are spread more widely over the phases. As the differences revolve around Grant, there does not seem any reason to attach any biological significance to these results; the relatively small sample size (nine) and random variation (narrow range of phases) seem to be more likely reasons.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.743	--	--	.972	.731	.999	.387	.835	.236
Dart			--	--	.723	.484	.954	.721	.984	.503
Grant					--	--	--	--	--	--
Lisbon							.672	.152	.722	.070
Pretoria									1.000	.606

Table A6.5a. Vault sutures, sexes pooled, age 20-29: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.981	.895	--	--	.526	.212	.526	.605	.947	.167
Dart			--	--	.755	.242	.576	.641	.873	.191
Grant					--	--	--	--	--	--
Lisbon							.794	.384	.947	.915
Pretoria									.530	.232

Table A6.5b. Vault sutures, sexes pooled, age 30-39: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.219	.037	.076	.026	.748	.375	.914	.335	1.000	.870
Dart			.005	**.001**	.007	**.001**	.972	.333	.438	.055
Grant					.212	.019	.033	.010	.068	.017
Lisbon							.109	.065	.812	.276
Pretoria									.990	.433

Table A6.5c. Vault sutures, sexes pooled, age 40-49: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.081	.021	--	.289	.854	.098	1.000	.624	1.000	.801
Dart			--	.157	.617	.227	.305	.058	.290	.067
Grant					--	.106	--	.258	--	.321
Lisbon							.854	.275	.987	.275
Pretoria									1.000	.852

Table A6.5d. Vault sutures, sexes pooled, age 50-59: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.806	.348	.998	.892	.704	.368	.188	.065	.985	.398
Dart			.958	.454	1.000	.841	.989	.272	.734	.070
Grant					.994	.523	.690	.187	.990	.724
Lisbon							.713	.183	.637	.082
Pretoria									.194	.009

Table A6.5e. Vault sutures, sexes pooled, age 60-69: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.819	.900	--	.099	1.000	1.000	.996	.725	.999	.362
Dart			--	.203	.987	.885	.993	.940	.336	.518
Grant					--	.126	--	.130	--	.142
Lisbon							1.000	.751	.928	.423
Pretoria									.653	.294

Table A6.5f. Vault sutures, sexes pooled, age 70-79: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.056	.013	.964	.173	1.000	.564	.597	.021	1.000	.820
Dart			1.000	.949	.122	.044	.892	.540	.099	.021
Grant					.979	.277	1.000	.730	.952	.165
Lisbon							.652	.070	.998	.468
Pretoria									.609	.028

Table A6.5g. Vault sutures, sexes pooled, age 80-89: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.675	.615	--	--	.955	.416	.635	.439	--	.709
Dart			--	--	.464	.568	.909	.340	--	.570
Grant					--	--	--	--	--	--
Lisbon							.816	.865	--	.815
Pretoria									--	1.000

Table A6.5h. Vault sutures, sexes pooled, age 90-99: two -sample K-S and MWU results

A.6.6 Fourth Rib Phase Distribution by Age Group (K-S and MWU Tests)

Finally, the fourth rib was also tested by age group for distribution and median variation (see Tables A6.6a to A6.6h). Grant did not have enough individuals with observable ribs in the 20 to 29, 30 to 39, 40 to 49, 60 to 69, 70 to 79, 80 to 89, or 90 to 99 age groups, and so was excluded.

No significant differences were found in any of the age groups. Near-significant differences in phase distribution and median were found between Coimbra and Lisbon for the 50 to 59 age group. Both collections had individuals in only narrow ranges of phases, and Lisbon's was higher than that of Coimbra. For the 60 to 69 age group, significant differences were found in phase distribution and median between Coimbra and Spitalfields, again the result of different and narrow ranges of phases for each collection. Other significant differences were in median only. As fourth rib sample sizes were small for all collections, meaningful interpretation of results is difficult.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.983	.555	--	--	.995	.511	.736	.308	1.000	.718
Dart			--	--	1.000	.968	1.000	.730	.998	.698
Grant					--	--	--	--	--	--
Lisbon							.997	.653	1.000	.692
Pretoria									.921	.398

Table A6.6a. Fourth rib, sexes pooled, age 20-29: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.961	.566	--	--	.734	.626	.999	.886	.566	.262
Dart			--	--	.375	.434	.847	.559	.576	.533
Grant					--	--	--	--	--	--
Lisbon							.925	.505	.242	.016
Pretoria									.637	.167

Table A6.6b. Fourth rib, sexes pooled, age 30-39: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.993	.688	--	--	.985	.346	1.000	.955	.530	.129
Dart			--	--	.994	.885	.990	.666	.988	.803
Grant					--	--	--	--	--	--
Lisbon							.988	.404	.925	.380
Pretoria									.465	.240

Table A6.6c. Fourth rib, sexes pooled, age 40-49: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.329	.452	.320	.068	.045	.007	.177	.053	.124	.013
Dart			.683	.693	.775	.778	.685	.642	.641	.774
Grant					.847	.217	.996	.858	.876	.548
Lisbon							.893	.203	.935	.918
Pretoria									.904	.355

Table A6.6d. Fourth rib, sexes pooled, age 50-59: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.847	.127	--	--	.280	.347	.593	.327	.022	.004
Dart			--	--	.867	.434	1.000	.544	.927	.401
Grant					--	--	--	--	--	--
Lisbon							.971	.742	.181	.063
Pretoria									.427	.053

Table A6.6e. Fourth rib, sexes pooled, age 60-69: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.974	.479	--	--	.978	.787	.998	.740	.998	.896
Dart			--	--	1.000	1.000	1.000	.662	1.000	.461
Grant					--	--	--	--	--	--
Lisbon							.996	.864	.996	.606
Pretoria									1.000	.744

Table A6.6f. Fourth rib, sexes pooled, age 70-79: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.998	1.000	--	--	.893	.623	1.000	.304	.988	.381
Dart			--	--	.518	.293	.586	.270	.593	.157
Grant					--	--	--	--	--	--
Lisbon							.441	.140	.976	.669
Pretoria									.512	.061

Table A6.6g. Fourth rib, sexes pooled, age 80-89: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.867	1.000	--	--	--	1.000	.964	.439	--	--
Dart			--	--	--	.083	.867	1.000	--	--
Grant					--	--	--	--	--	--
Lisbon							--	.480	--	--
Pretoria									--	--

Table A6.6h. Fourth rib, sexes pooled, age 90-99: two -sample K-S and MWU results

Appendix 7: Buckberry-Chamberlain Scored Traits Score Distribution by Age Group

A.7.1 Transverse Organisation Score Distribution by Age Group (K-S and MWU Tests)

While some significant differences in score distribution and median were found between collections (for example, for transverse organisation, between Lisbon and Dart for the 20 to 29 year olds), they were largely in median only. For the 60 to 69 age group, significant differences were found between Grant and the other collections. The Grant individuals aged 60 to 69 differ in that there were more scores of 5, while other collections tended to have lower scores for this age group. However, the differences did not remain for the older age groups. The high Grant scores for ages 60 to 69 may be stochastic. There were not many differences, though, and no particular patterns can be seen. Tables A7.1a to A7.1h give all K-S and MWU test results for transverse organisation; significant differences are in bold, slightly larger type.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.339	.024	.538	.368	.866	.302	.903	.305	.975	.911
Dart			.985	.870	.013	**.003**	.590	.116	.244	.098
Grant					.660	.176	.523	.563	.980	.354
Lisbon							.117	.044	.999	.581
Pretoria									.734	.441

Table A7.1a. Buckberry-Chamberlain transverse organisation scores, age 20-29: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.999	.635	.916	.479	.865	.159	.978	.328	.434	.269
Dart			.433	.266	.972	.443	.999	.720	.972	.414
Grant					.259	.038	.507	.092	.072	.158
Lisbon							1.000	.627	.794	.804
Pretoria									.914	.594

Table A7.1b. Buckberry-Chamberlain transverse organisation scores, age 30-39: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.914	.213	.586	.051	1.000	.870	.875	.606	.914	.182
Dart			.999	.268	.794	.152	.152	.037	.972	.721
Grant					.531	.033	.087	.006	.997	.530
Lisbon							.794	.766	.794	.134
Pretoria									.152	.052

Table A7.1c. Buckberry-Chamberlain transverse organisation scores, age 40-49: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.329	.086	.565	.141	.329	.158	.714	.173	.599	.354
Dart			.997	.752	1.000	.772	1.000	.740	.368	.033
Grant					1.000	.946	.995	.962	.376	.076
Lisbon							1.000	.964	.584	.063
Pretoria									.393	.061

Table A7.1d. Buckberry-Chamberlain transverse organisation scores, age 50-59: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.169	.022	.013	**.002**	.507	.192	.194	.062	1.000	.644
Dart			**.003**	**.000**	1.000	.479	.684	.797	.453	.057
Grant					.016	**.001**	.013	**.000**	.011	**.002**
Lisbon							.872	.420	.913	.331
Pretoria									.518	.098

Table A7.1e. Buckberry-Chamberlain transverse organisation scores, age 60-69: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.998	.256	.765	.155	.998	.256	1.000	.504	1.000	.560
Dart			.244	.015	1.000	1.000	1.000	.631	.911	.083
Grant					.244	.015	.423	.043	.982	.369
Lisbon							1.000	.631	.911	.083
Pretoria									.996	.207

Table A7.1f. Buckberry-Chamberlain transverse organisation scores, age 70-79: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.402	1.000	.569	.988	.216	.475	.154	.948	.134
Dart			1.000	.916	.560	.029	.132	.031	1.000	.413
Grant					.888	.105	.398	.108	1.000	.436
Lisbon							.930	.619	.334	.008
Pretoria									.088	.017

Table A7.1g. Buckberry-Chamberlain transverse organisation scores, age 80-89: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.735	.441	.114	.985	.322	1.000	.903	--	.617
Dart			.230	.041	.983	.304	1.000	.883	--	--
Grant					.203	.060	.518	.144	--	--
Lisbon							1.000	.431	--	--
Pretoria									--	--

Table A7.1h. Buckberry-Chamberlain transverse organisation scores, age 90-99: two -sample K-S and MWU results

A.7.2 Surface Texture Score Distribution by Age Group (K-S and MWU Tests)

Significant and near-significant differences in score distribution and median are found throughout the age groups. From the 40 to 49 age group onwards, Spitalfields had significantly different or near-significantly different medians (and sometimes score distributions) compared to Dart and Pretoria. These differences reflect Spitalfields higher scores in general from age 40 and older. The Grant Collection had significant or near-significant differences compared to the other collections (with the exception of Spitalfields, for most age groups). This reflects Grant's relatively high scores compared to the other collections. For individuals aged 20 to 29 years, the significant differences also reflect Grant's low numbers of individuals in this age group. Tables A7.2a to A7.2h give the K-S and MWU results.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.004	**.000**	.047	**.004**	.978	.298	.172	.005	.047	.014
Dart			.106	.058	.013	**.001**	.338	.046	.942	.661
Grant					.047	.006	.047	.017	.227	.071
Lisbon							.819	.048	.097	.049
Pretoria									.761	.384

Table A7.2a. Buckberry-Chamberlain surface texture scores, age 20-29: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.799	.217	.029	.981	.463	.676	.299	1.000	.619
Dart			.246	.046	.972	.309	.987	.535	1.000	.816
Grant					.032	.004	.097	.044	.415	.093
Lisbon							.566	.044	.794	.235
Pretoria									.839	.632

Table A7.2b. Buckberry-Chamberlain surface texture scores, age 30-39: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.239	.044	.010	**.001**	.956	.244	.621	.215	.005	**.000**
Dart			.035	**.005**	.968	.447	1.000	.295	.024	.013
Grant					.046	**.003**	.010	**.001**	.998	.296
Lisbon							1.000	.901	.034	**.005**
Pretoria									.005	**.001**

Table A7.2c. Buckberry-Chamberlain surface texture scores, age 40-49: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.483	.118	.857	.140	1.000	.474	.938	.328	.540	.048
Dart			.063	.006	.137	.046	1.000	.542	.011	**.001**
Grant					1.000	.491	.213	.024	1.000	.781
Lisbon							.460	.129	.964	.290
Pretoria									.061	.005

Table A7.2d. Buckberry-Chamberlain surface texture scores, age 50-59: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.492	.053	.007	**.000**	1.000	.755	.123	.020	.928	.065
Dart			.010	**.000**	.914	.120	.954	.912	.055	**.002**
Grant					.023	**.001**	.004	**.000**	.107	.019
Lisbon							.420	.066	.560	.060
Pretoria									.006	**.000**

Table A7.2e. Buckberry-Chamberlain surface texture scores, age 60-69: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.209	.049	.996	.158	1.000	.926	.738	.127	.940	.091
Dart			.032	**.004**	.123	.036	.999	.548	.018	**.002**
Grant					.883	.092	.199	.008	1.000	.805
Lisbon							.560	.107	.689	.046
Pretoria									.136	**.003**

Table A7.2f. Buckberry-Chamberlain surface texture scores, age 70-79: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.573	.071	.792	.041	.990	.350	.886	.234	.969	.161
Dart			.162	**.002**	.999	.363	1.000	.548	.266	.007
Grant					.486	.011	.320	.008	1.000	.519
Lisbon							1.000	.771	.726	.039
Pretoria									.508	.027

Table A7.2g. Buckberry-Chamberlain surface texture scores, age 80-89: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.822	.636	.139	.025	.985	.260	.236	.046	--	1.000
Dart			.230	.064	.983	.741	.816	.151	--	--
Grant					.089	.023	.100	.031	--	--
Lisbon							.736	.187	--	--
Pretoria									--	--

Table A7.2h. Buckberry-Chamberlain surface texture scores, age 90-99: two -sample K-S and MWU results

A.7.3 Microporosity Score Distribution by Age Group (K-S and MWU Tests)

Significant differences were somewhat sporadic, but when they do appear are largely between Spitalfields compared to Dart and Pretoria, sometimes between Coimbra compared to Pretoria and Dart, and sometimes between Grant compared to Lisbon. These differences reflect the lower scores of Pretoria and Dart compared to Spitalfields and Coimbra. Lisbon scores tended to be lower than those of Grant. However, no age-related trends are seen (see Tables A7.3a to A7.3h).

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.283	.091	.621	.352	.408	.042	.722	.072	.585	.254
Dart			.342	.156	1.000	.866	.948	.768	.203	.029
Grant					.181	.107	.128	.100	1.000	.790
Lisbon							.993	.595	.066	.013
Pretoria									.036	.015

Table A7.3a. Buckberry-Chamberlain microporosity scores, age 20-29: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.242	.027	1.000	.746	.117	.016	.005	**.001**	1.000	.733
Dart			.531	.177	1.000	.963	.728	.375	.526	.041
Grant					.331	.139	.039	.029	.975	.936
Lisbon							.945	.361	.300	.025
Pretoria									.021	**.001**

Table A7.3b. Buckberry-Chamberlain microporosity scores, age 30-39: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.728	.096	.952	.251	.459	.065	.728	.124	1.000	.800
Dart			.997	.675	1.000	1.000	1.000	.791	.526	.078
Grant					1.000	.636	1.000	.832	.837	.196
Lisbon							1.000	.765	.300	.053
Pretoria									.526	.097

Table A7.3c. Buckberry-Chamberlain microporosity scores, age 40-49: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.035	**.002**	.488	.025	.978	.150	**.003**	**.000**	1.000	.459
Dart			.997	.329	.329	.050	.996	.333	.136	.012
Grant					.991	.386	.699	.075	.814	.128
Lisbon							.053	.006	1.000	.475
Pretoria									.017	**.001**

Table A7.3d. Buckberry-Chamberlain microporosity scores, age 50-59: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.341	.025	.835	.098	1.000	.679	.358	.039	.962	.230
Dart			1.000	.698	.211	.012	.929	.590	.738	.193
Grant					.648	.053	.974	.985	.844	.470
Lisbon							.219	.018	.819	.118
Pretoria									.996	.352

Table A7.3e. Buckberry-Chamberlain microporosity scores, age 60-69: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.047	**.003**	.859	.087	1.000	.617	.047	**.003**	1.000	.967
Dart			.739	.267	.082	.006	1.000	.880	.056	**.004**
Grant					.971	.170	.739	.300	.881	.100
Lisbon							.082	.006	1.000	.652
Pretoria									.056	**.004**

Table A7.3f. Buckberry-Chamberlain microporosity scores, age 70-79: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.996	.233	1.000	.453	1.000	.413	.875	.117	1.000	.407
Dart			.965	.157	1.000	.637	1.000	.660	.939	.119
Grant					.999	.228	.791	.096	1.000	1.000
Lisbon							1.000	.357	.997	.184
Pretoria									.720	.068

Table A7.3g. Buckberry-Chamberlain microporosity scores, age 80-89: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.822	.175	.893	.157	1.000	1.000	.952	.224	--	1.000
Dart			1.000	.780	.591	.079	1.000	.961	--	--
Grant					.832	.061	1.000	.847	--	--
Lisbon							.865	.112	--	--
Pretoria									--	--

Table A7.3h. Buckberry-Chamberlain microporosity scores, age 90-99: two -sample K-S and MWU results

A.7.4 Macroporosity Score Distribution by Age Group (K-S and MWU Tests)

While there were a few significant differences in macroporosity, they were largely in median only. No trends in differences between specific collections or at particular ages can be seen (Tables A7.4a to A7.4h).

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.989	.363	.987	.561	.211	.018	.238	.022	.999	.448
Dart			.985	.301	.819	.118	.854	.137	1.000	.963
Grant					.660	.042	.682	.049	.991	.345
Lisbon							1.000	.957	.886	.137
Pretoria									.910	.156

Table A7.4a. Buckberry-Chamberlain macroporosity scores, age 20-29: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.197	.018	.979	.467	.377	.051	.172	.014	.995	.395
Dart			.990	.219	1.000	.636	1.000	.957	.794	.116
Grant					1.000	.404	.984	.196	1.000	.936
Lisbon							1.000	.594	.972	.259
Pretoria									.755	.100

Table A7.4b. Buckberry-Chamberlain macroporosity scores, age 30-39: two -sample K-S and MWU results

The differences in macroporosity score distribution between Lisbon and Coimbra 30 to 39 year olds approaches significance, with a p-value of 0.051.

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.963	.236	.095	.399	.119	.997	.745	.905	.337
Dart			.642	.139	.794	.184	1.000	.808	1.000	.429
Grant					.164	.015	.642	.106	.259	.035
Lisbon							.972	.289	1.000	.544
Pretoria									1.000	.604

Table A7.4c. Buckberry-Chamberlain macroporosity scores, age 40-49: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.819	.436	.375	.038	.819	.313	.714	.402	1.000	.456
Dart			.925	.264	1.000	.872	1.000	.940	.497	.183
Grant					.996	.309	.900	.313	.195	**.012**
Lisbon							1.000	.939	.497	.115
Pretoria									.406	.173

Table A7.4d. Buckberry-Chamberlain macroporosity scores, age 50-59: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.765	.549	.612	.757	.735	.245	.358	.171	.872	.167
Dart			.999	.525	.999	.724	1.000	.575	.096	.111
Grant					.648	.307	.835	.246	.336	.617
Lisbon							1.000	.794	.082	.018
Pretoria									.021	.014

Table A7.4e. Buckberry-Chamberlain macroporosity scores, age 60-69: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.941	.221	.996	.444	1.000	1.000	.998	.306	1.000	.814
Dart			1.000	.623	.978	.224	1.000	.802	.791	.265
Grant					.971	.451	1.000	.798	1.000	.540
Lisbon							.978	.310	.999	.817
Pretoria									.964	.374

Table A7.4f. Buckberry-Chamberlain macroporosity scores, age 70-79: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.305	.079	.886	.319	.948	.220	1.000	.569	.501	.248
Dart			.413	.036	.978	.485	.631	.195	.134	.028
Grant					.413	.068	.673	.165	1.000	.862
Lisbon							1.000	.499	.134	.053
Pretoria									.286	.127

Table A7.4g. Buckberry-Chamberlain macroporosity scores, age 80-89: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.617	.163	1.000	.480	.985	.441	.388	.055	--	.221
Dart			.617	.162	1.000	.514	.936	.477	--	--
Grant					.938	.294	.249	.078	--	--
Lisbon							.665	.223	--	--
Pretoria									--	--

Table A7.4h. Buckberry-Chamberlain macroporosity scores, age 90-99: two -sample K-S and MWU results

There was only one individual from Spitalfields in the 90 to 99 age group, so the K-S test could not be done with Spitalfields.

A.7.5 Apical Change Score Distribution by Age Group (K-S and MWU Tests)

As with macroporosity, there were some significant differences between collections in terms of apical change, but they were mostly in median only. No patterns emerge in terms of differences between specific combinations of collections, or in particular age groups (Tables A7.5a to A7.5h).

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.841	.381	.048	.741	.123	.484	.050	.734	.176
Dart			.375	.055	.560	.080	.329	.031	.858	.234
Grant					.181	.006	.120	**.002**	.429	.148
Lisbon							1.000	.637	.066	.006
Pretoria									.030	**.002**

Table A7.5a. Buckberry-Chamberlain apical change scores, age 20-29: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.815	.995	.395	1.000	.962	.437	.084	1.000	.459
Dart			.998	.564	1.000	.854	.434	.070	1.000	.634
Grant					.998	.424	.150	.024	1.000	.921
Lisbon							.434	.083	1.000	.495
Pretoria									.234	.028

Table A7.5b. Buckberry-Chamberlain apical change scores, age 30-39: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.938	.245	.952	.327	1.000	.846	1.000	.639	1.000	.568
Dart			.818	.089	.972	.212	.993	.432	.526	.083
Grant					.999	.444	.781	.180	.965	.513
Lisbon							1.000	.527	1.000	.754
Pretoria									.997	.273

Table A7.5c. Buckberry-Chamberlain apical change scores, age 40-49: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.598	.747	.175	.978	.580	.226	.063	1.000	.801
Dart			.526	.084	.819	.225	.415	.147	1.000	.801
Grant					.526	.192	.234	.016	.680	.138
Lisbon							.019	.009	.937	.393
Pretoria									.367	.115

Table A7.5d. Buckberry-Chamberlain apical change scores, age 50-59: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.667	.308	.034	.841	.700	1.000	1.000	1.000	.843
Dart			.292	.024	.804	.503	1.000	.703	1.000	.828
Grant					.728	.184	.564	.057	.363	.034
Lisbon							.996	.723	.900	.611
Pretoria									1.000	.860

Table A7.5e. Buckberry-Chamberlain apical change scores, age 60-69: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.996	.291	.999	.419	1.000	.743	.681	.143	1.000	.552
Dart			.576	.075	1.000	.465	1.000	.794	.804	.100
Grant					.971	.267	.199	.024	1.000	.799
Lisbon							.819	.277	1.000	.357
Pretoria									.334	.033

Table A7.5f. Buckberry-Chamberlain apical change scores, age 70-79: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	.996	.256	.394	.074	1.000	.941	.954	.286	.644	.144
Dart			.071	.010	.978	.203	1.000	.812	.150	.021
Grant					.388	.073	.046	.007	1.000	.703
Lisbon							.886	.230	.646	.144
Pretoria									.099	.015

Table A7.5g. Buckberry-Chamberlain apical change scores, age 80-89: two -sample K-S and MWU results

	Dart		Grant		Lisbon		Pretoria		Spitalfields	
	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU	K-S	MWU
Coimbra	1.000	.891	1.000	.576	1.000	.673	.998	.453	--	.617
Dart			1.000	.554	1.000	.480	.990	.370	--	.546
Grant					.989	.312	1.000	1.000	--	.480
Lisbon							.804	.181	--	.705
Pretoria									--	.386

Table A7.5h. Buckberry-Chamberlain apical change scores, age 90-99: two -sample K-S and MWU results

While Grant has been included, there were only two individuals in the 90 to 99 age category. Kolomogorov-Smirnov tests have not been done for Spitalfields, as there was only one individual in this age category.

Appendix 8: Mean Age per Phase/Score

A.8.1 Lateral-Anterior Suture Closure Mean Age by Phase

No significant differences in the female data were found for mean age per phase for lateral-anterior suture closure phases (Table A8.1).

	Phase					
	3	4	5	6	7	8
Levene's test	.114	--	.217	.326	.742	--
ANOVA	.293*/.300***	.567***	.966	.119	.602	--
RTEM	--	--	--	--	--	--

Table A8.1. Lateral-anterior sutures, females only: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case; ***t-test, only two collections; -- : n/a

A.8.2 Vault Cranial Suture Closure Mean Age by Phase

Few significant differences were found in mean age per phase for vault suture closure phases. Results are presented in Tables A8.2 to A8.4 for the sexes pooled, females only and males only.

	Phase					
	2	3	4	5	6	7
Levene's test	.700	.246	.093	**.027***/.059**	.135	.121
ANOVA	.614	.325	.206*/.251**	.684*/.731**	.580	.584
RTEM	--	--	--	.180	--	--

Table A8.2. Vault suture phases, sexes pooled: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case; ***t-test, only two collections; -- : n/a

	Phase					
	2	3	4	5	6	7
Levene's test	.483	.180	.191	.425	.899	.141
ANOVA	.592	.391	.535	.985*/.975**	**.017**	.707
RTEM	--	--	--	--	**--**	--

Table A8.3. Vault sutures, females only: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case; ***t-test, only two collections; -- : n/a

	Phase					
	2	3	4	5	6	7
Levene's test	--	**.001**	.389	**.038**	.372	.845
ANOVA	--	.281*/.182**	.063*/.065**	.479	.467	.738
RTEM	--	.400	--	.302	--	--

Table A8.4. Vault sutures, males only: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case; ***t-test, only two collections; -- : n/a

A.8.3 Sternal End of Fourth Rib Mean Age by Phase

Few significant differences in mean were found in the fourth rib phases. None were found for the sexes pooled (Table A8.5), while only phase 10 had significant differences for both females only and males only (Tables A8.5 and A8.7).

	Phase									
	5	6	7	8	9	10	11	12	13	14
Levene's test	**.033**	**.003**	.280	.419	.550	.196	**.021**	.244	--	.296
ANOVA	.566	.110	.482	.164	.171* /.125**	.607*/ .482**	.323	.629	.356*/ .959***	.900
RTEM	.710	.086	--	--	--	--	.360	--	--	--

Table A8.5. Fourth rib phases, sexes pooled: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case; ***t-test, only two collections; -- : n/a

	Phase									
	5	6	7	8	9	10	11	12	13	14
Levene's test	.129	**.001**	**.022**	.264	.296	.095	.059	--	--	--
ANOVA	.795*/ .683**	.134	.437	.502	.181	**.048***/ **.027****	.840*/ .933***	--	--	--
RTEM	--	.103	.561	--	--	--	--	--	--	--

Table A8.6. Fourth rib phases, females only: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case; ***t-test, only two collections; -- : n/a

	Phase									
	5	6	7	8	9	10	11	12	13	14
Levene's test	--	**.001**	.237	.187	.310	.415	.074	.360	--	.296
ANOVA	--	.600*/ .557***	.929	.182	.566*/ .682**	**.050***/ **.029****	.297	.571	.356*/ .959***	.900
RTEM	--	.557***	--	--	--	--	--	--	--	--

Table A8.7. Fourth rib phases, males only: ANOVA and test of variance
RTEM: included when necessary; *including all collections; **excluding those with only one case; ***t-test, only two collections; -- : n/a

A.8.4 Buckberry-Chamberlain Trait Scores

A.8.4.1 Transverse Organisation Mean Age by Score

The only significant difference in mean for transverse organisation was found in the males-only data, for scores of 5.

	Score				
	1	2	3	4	5
Levene's test	--	**.014**	.239	.060	.645
ANOVA	--	.117/.111*	.219	.483	.089

Table A8.8. Buckberry-Chamberlain transverse organisation score, sexes pooled: ANOVA and test of variance
*RTEM

	Score				
	1	2	3	4	5
Levene's test	--	**.001**	.267	.677	.510
ANOVA	--	.061/.071*	.406	.199	.647

Table A8.9. Buckberry-Chamberlain transverse organisation score, females only: ANOVA and test of variance
*RTEM

	Score				
	1	2	3	4	5
Levene's test	--	.250	.878	.100	.550
ANOVA	--	.853	.349	.804	**.013**

Table A8.10. Buckberry-Chamberlain transverse organisation score, males only: ANOVA and test of variance
*RTEM

A.8.4.2 Surface Texture Mean Age by Score

For surface texture scores, significant differences for the sexes pooled were found in scores of 3 and 4 (Table A8.11). No significant differences were found for females only (Table A8.12), but for males only, a significant difference in mean for scores of 3 was found (Table A8.13).

	Score				
	1	2	3	4	5
Levene's test	.432	**.006**	**.011**	**.046**	.482
ANOVA	.821	.696/.716*	.076/**.044***	.002/**.001***	.485

Table A8.11. Buckberry-Chamberlain surface texture score, sexes pooled: ANOVA and test of variance
*RTEM

	Score				
	1	2	3	4	5
Levene's test	.096	.052	.579	.143	.864
ANOVA	.732	.528	.508	.215	.234

Table A8.12. Buckberry-Chamberlain surface texture score, females only: ANOVA and test of variance
*RTEM

	Score				
	1	2	3	4	5
Levene's test	.383	**.049**	**.003**	.307	.613
ANOVA	.166	.958/.941*	.081/**.001***	.078	.965

Table A8.13. Buckberry-Chamberlain surface texture score, males only: ANOVA and test of variance
*RTEM

A.8.4.3 Microporosity Mean Age by Score

Significant differences in mean ages for microporosity scores were found for the sexes pooled, for all scores (1, 2 and 3) (Table A8.14). No significant differences in mean for the females-only scores were found (Table A8.15). Significant differences in the males-only means were found for scores of 1 and 2 (Table A8.16).

	Score		
	1	2	3
Levene's test	**.015**	**.031**	.232
ANOVA	.014/**.017***	.005/**.005***	**.010**

Table A8.14. Buckberry-Chamberlain microporosity score, sexes pooled: ANOVA and test of variance
*RTEM

	Score		
	1	2	3
Levene's test	.450	.905	.905
ANOVA	.356	.151	.174

Table A8.15. Buckberry-Chamberlain microporosity score, females only: ANOVA and test of variance

*RTEM

	Score		
	1	2	3
Levene's test	**.005**	**.014**	.220
ANOVA	.032/**.001***	.017/**.045***	.113

Table A8.16. Buckberry-Chamberlain microporosity score, males only: ANOVA and test of variance

*RTEM

A.8.4.4 Macroporosity Mean Age by Score

For macroporosity, a significant difference in mean was found in scores of 2 for the sexes pooled. No other significant differences in mean were found, for the sexes pooled, the females only, or the males only (see Tables A8.17 to A8.19).

	Score		
	1	2	3
Levene's test	**.001**	.127	.425
ANOVA	.043/.056*****	**.033**	.593

Table A8.17. Buckberry-Chamberlain macroporosity score, sexes pooled: ANOVA and test of variance

*RTEM

	Score		
	1	2	3
Levene's test	.129	.161	.671
ANOVA	.403	.232	.590

Table A8.18. Buckberry-Chamberlain macroporosity score, females only: ANOVA and test of variance

	Score		
	1	2	3
Levene's test	**.004**	.627	.612
ANOVA	.111/.207*	.262	.975

Table A8.19. Buckberry-Chamberlain macroporosity score, males only: ANOVA and test of variance

*RTEM

A.8.4.5 Apical Change Mean Age by Score

For apical change, a significant difference in mean was found in scores of 2 for the sexes pooled. No other significant differences were found for apical change score means (see Tables A8.20 to A8.22).

	Score		
	1	2	3
Levene's test	.634	.235	.901
ANOVA	.770	**.023**	.152

Table A8.20. Buckberry-Chamberlain apical change score, sexes pooled: ANOVA and test of variance

	Score		
	1	2	3
Levene's test	.292	.440	.694
ANOVA	.305	.052	.716

Table A8.21. Buckberry-Chamberlain apical change score, females only: ANOVA and test of variance

	Score		
	1	2	3
Levene's test	.397	.364	.768
ANOVA	.664	.250	.131

Table A8.22. Buckberry-Chamberlain apical change score, males only: ANOVA and test of variance

Appendix 9: Overall vs Subjective Age Estimates by Sex

A.9.1 Grant Collection

The Grant Collection overall estimates were low throughout the age groups (see Table A9.1), although for males only and the pooled results, the lowest percentages of correct estimates were in the oldest age groups (from the 70 to 79 age group and older). The female age estimates were incorrect in all individuals for a number of age categories, so there were no clear patterns. The female sample is so small (*n* = 18) that it would not be meaningful to compare with the male estimates.

Known Age Group	Number of Individuals			Correct Age Estimates (n)			Correct Age Estimates (%)		
	F	M	**P**	F	M	**P**	F	M	**P**
20-29	2	3	**5**	1	1	**2**	50.0	33.3	**40.0**
30-39	1	10	**11**	0	6	**6**	0.0	60.0	**54.6**
40-49	2	9	**11**	0	3	**3**	0.0	33.3	**27.3**
50-59	--	11	**--**	--	10	**--**	--	90.9	**--**
60-69	5	10	**15**	4	4	**8**	80.00	40.0	**53.3**
70-79	5	13	**18**	0	3	**3**	0.0	23.1	**16.7**
80-89	1	9	**10**	0	0	**0**	0.0	0.0	**0.0**
90-99	2	--	**--**	0	--	**--**	0.0	--	**--**
100+	--	--	**--**	--	--	**--**	--	--	**--**
Total	18	65	**83**	5	27	**32**	27.8	41.5	**38.6**

Table A9.1. Grant Collection, numbers and percentages of correct overall estimates
F: female; M: male; P: pooled (sexes); --: n/a.

Known Age Group	Number of Individuals			Correct Age Estimates (n)			Correct Age Estimates (%)		
	F	M	**P**	F	M	**P**	F	M	**P**
20-29	2	3	**5**	0	0	**0**	0.0	0.0	**0.0**
30-39	1	10	**11**	0	9	**9**	0.0	90.0	**81.8**
40-49	2	9	**11**	1	4	**5**	50.0	44.4	**45.4**
50-59	--	11	**--**	--	5	**--**	--	45.4	**--**
60-69	5	10	**15**	0	2	**2**	0.0	20.0	**13.3**
70-79	5	13	**18**	1	1	**2**	20.0	7.7	**11.1**
80-89	1	9	**10**	0	1	**1**	0.0	11.1	**10.0**
90-99	2	--	**--**	0	--	**--**	0.0	--	**--**
100+	--	--	**--**	--	--	**--**	--	--	**--**
Total	18	65	**83**	2	22	**24**	11.1	33.8	**28.9**

Table A9.2. Grant Collection, numbers and percentages of correct subjective estimates
F: female; M: male; P: pooled (sexes); --: n/a.
Inaccuracy, bias and standard deviation for the sexes pooled are discussed in Chapter 4.

As with the overall estimates, there was a decrease in correct subjective estimates (Table A9.2) in the older age groups for males and the pooled estimates (from the 60 to 69 group and older). The females, again, had too small a sample size and too many incorrect estimates throughout the age groups to show any clear patterns.

For the females only (Table A9.3), the patterns were somewhat obscured again by small sample size; however, inaccuracy, standard deviation and bias again increased with age, beginning from the 70 to 79 age group. Bias again was positive for the younger age groups and negative for the older age groups (from 60 to 69 for overall estimates and 70 to 79 for subjective estimates). The total (mean) inaccuracy was slightly greater for the subjective estimates compared to the overall estimates, while total standard deviation was slightly lower for the subjective estimates. The total bias was the same for overall and subjective estimates. The pattern of bias generally followed the parabolic pattern, but some age groups had only one individual – means could not be calculated for these, again making it difficult to see if the pattern held.

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	3.00	4.00	4.24	1.41	3.00	4.00
30-39	6.0*	9.0*	--	--	6.0*	9.0*
40-49	10.00	1.50	7.07	2.12	10.00	1.50
50-59	--	--	--	--	--	--
60-69	1.40	4.80	3.13	3.42	-1.40	0.40
70-79	6.80	8.40	6.76	7.13	-6.80	-8.40
80-89	10.0*	3.0*	--	--	-10.0*	-3.0*
90-99	12.50	10.50	0.71	2.12	-12.50	-10.50
100+	--	--	--	--	--	--
Total (mean)	6.00	6.11	5.69	4.87	-2.44	-2.44

Table A9.3. Grant Collection, inaccuracy, standard deviation and bias for each age group, females only
*not a mean, based on one individual only.

For males (Table A9.4), inaccuracy began to increase steadily for both the overall and subjective estimates at the 60 to 69 age group, as did standard deviation for the overall estimates. The subjective estimates increased in standard deviation from the youngest to the oldest age group. Bias followed the parabolic pattern (again with a minor fluctuation for overall estimates at the 30 to 39 and 40 to 49 age groups) with underageing beginning at the 60 to 69 and 40 to 49 age groups for the overall and subjective estimates, respectively. Total mean inaccuracy, standard deviation and bias were all greater for the subjective estimates than the overall estimates.

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	7.00	4.33	8.89	0.58	7.00	4.33
30-39	2.20	0.40	3.19	1.26	2.20	0.40
40-49	3.11	4.22	3.55	4.09	2.44	-2.44
50-59	0.55	4.27	1.81	5.71	0.55	-4.27
60-69	5.00	7.10	5.35	6.23	-5.00	-7.10
70-79	10.00	8.54	7.87	6.10	-10.00	-7.62
80-89	16.00	20.11	8.20	9.36	-16.00	-20.11
90-99	--	--	--	--	--	--
100+	--	--	--	--	--	--
Total (mean)	6.17	7.15	7.49	8.06	-3.89	-6.2

Table A9.4. Grant Collection, inaccuracy, standard deviation and bias for each age group, males only

A.9.2 Spitalfields Collection

The pattern of correct overall age estimates by age group mirrored that of bias (the parabolic pattern), with a low initial percentage of correct estimates for the youngest age group, which increased steadily (here, the highest percentage for overall estimates for males, females and the sexes pooled was the 40 to 49 age group) until a reversal in the middle to older age groups, decreasing in percentage after that. The decrease began for the overall estimates at the 50 to 59 age group. The male and female percentages were quite close from the 20 to 29 age group until the 50 to 59 age group; the largest difference occurred in the 60 to 69 group, where 40.0% of females were aged correctly compared to only 18.9% of males. After that group, no more individuals were aged correctly. Full details are in Table A9.5, below.

Known Age Group	Number of Individuals			Correct Age Estimates (n)			Correct Age Estimates (%)		
	F	M	P	F	M	P	F	M	P
20-29	8	6	14	3	2	5	37.5	33.3	35.7
30-39	10	11	21	7	8	15	70.0	72.7	71.4
40-49	10	10	20	8	9	17	80.0	90.0	85.0
50-59	10	9	19	6	5	11	60.0	55.6	57.9
60-69	10	11	21	4	2	6	40.0	18.9	28.6
70-79	10	8	18	0	0	0	0.0	0.0	0.0
80-89	10	3	13	0	0	0	0.0	0.0	0.0
90-99	--	1	--	--	0	--	--	0.0	--
100+									
Total	68	59	127	28	26	54	41.2	44.0	42.5

Table A9.5. Spitalfields Collection, numbers and percentages of correct overall estimates
F: female; M: male; P: pooled (sexes); --: n/a.

The subjective Spitalfields estimates followed the same pattern of increasing percentages of correct estimates until a reversal, then decreasing percentages with age – interestingly, until the 80 to 89 age group, which had higher percentages of correct estimates than the two preceding age groups. For females, correct estimates increased in percentage only until the 30 to 39 age group, and then decreased from the 40 to 49 group steadily (until 80 to 89). Percentages of correct male estimates increased until the 40 to 49 age group, then decreased from the 50 to 59 group until ages 80 to 89. For the sexes pooled, the reversal occurred after the 30 to 39 age group, with decreasing percentages of correct estimates thereafter, with the exception of the 80 to 89 age group. Male and female values were not very different; the greatest differences were in the 30 to 39 and 80 to 89 age groups, but in terms of absolute numbers, the difference was not great. For the 30 to 39 group (where females were aged correctly in 90.0% of cases, and males in 72.7% of cases), one female was incorrectly aged, while two males were incorrectly aged. For the 80 to 89 age group (80.0% of females were aged correctly compared to 60.0% of males), two males and two females were incorrectly aged (out of ten females and five males in that age group). There was only one individual (male) in the 90 to 99 age group, incorrectly aged. Details are in Table A9.6, below.

Known Age Group	Number of Individuals			Correct Age Estimates (n)			Correct Age Estimates (%)		
	F	M	P	F	M	P	F	M	P
20-29	9	6	15	6	4	10	66.7	66.7	66.7
30-39	10	11	21	9	8	17	90.0	72.7	81.0
40-49	10	10	20	7	8	5	70.0	80.0	75.0
50-59	10	10	20	6	6	12	60.0	60.0	60.0
60-69	10	11	21	5	6	11	50.0	54.6	52.4
70-79	10	11	21	5	5	10	50.0	45.4	47.6
80-89	10	5	15	8	3	11	80.0	60.0	73.3
90-99	-	1	--	--	0	--	--	0.0	--
100+	--	--	--	--	--	--	--	--	--
Total	69	65	134	46	40	86	66.7	61.5	64.2

Table A9.6. Spitalfields Collection, numbers and percentages of correct subjective estimates
F: female; M: male; P: pooled (sexes); --: n/a.

The patterns of inaccuracy, standard deviation, and bias in the overall estimates for both females only and males only were the same as that of the sexes pooled. The only exception was a slight decrease in female standard deviation in the 70 to 79 group; otherwise the values all began high, decreased until the 40 to 49 group, and increased from the 50 to 59 group, after which values rose with age. The fluctuation in bias at the 40 to 49 and 50 to 59 age groups were also present in both males and females. Table A9.7 has full details for females only and Table A9.8 for males only. The subjective inaccuracy and standard deviation for females and males fluctuated as did the values for the sexes pooled. The female values for subjective estimate bias followed the same pattern of increasing values as did the values for the sexes pooled, with the same decrease in bias in the 80 to 89 age group. However, the bias in the 20 to 29 year old females was negative, while that of the 30 to 39 through to the 50 to 59 year olds was positive (as with the sexes pooled), before switching to negative bias in the 60 to 69 age group, again, as with the sexes pooled. The male pattern of bias was not the same, and instead fluctuates across the age groups. The bias was positive from the 20 to 29 year olds through to the 50 to 59 year olds, changing to negative bias in the 60 to 69 age group, which then rose in value with age, again with the exception of the 80 to 89 age group. There was one male in the 90 to 99 age group, however, with a highly negative bias. For both males and females, inaccuracy, standard deviation and bias were lower in the subjective estimates compared to the overall estimates. For both overall and subjective bias, the values for females were higher than those for males – overall female bias is -3.19, while overall male bias is -1.27; subjective female bias is -1.13, while subjective male bias is -0.25.

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	6.75	0.67	7.27	1.12	6.75	-0.44
30-39	1.90	0.10	3.75	0.32	1.90	0.10
40-49	1.10	1.60	2.60	2.67	-0.50	0.60
50-59	1.90	3.40	3.11	4.74	0.30	1.20
60-69	5.20	2.80	6.29	3.58	-5.20	-2.60
70-79	7.90	4.60	5.86	11.19	-7.90	-4.40
80-89	15.70	2.30	7.06	4.85	-15.70	-2.30
90-99	--	--	--	--	--	--
100+	--	--	--	--	--	--
Total (mean)	5.75	2.23	7.04	5.25	-3.19	-1.13

Table A9.7. Spitalfields Collection, inaccuracy, standard deviation and bias for each age group, females only

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	9.33	1.33	10.15	2.80	9.33	1.33
30-39	4.00	0.82	7.28	1.66	4.00	0.82
40-49	0.30	1.70	0.95	5.03	-0.30	1.50
50-59	2.11	3.20	2.98	5.35	0.56	1.60
60-69	4.18	1.82	3.28	2.60	-3.82	-1.09
70-79	8.63	3.64	6.12	5.28	-8.63	-2.73
80-89	15.00	1.20	10.15	1.64	-15.00	-1.20
90-99	21*	16*	--	--	-21*	-16*
100+	--	--	--	--	--	--
Total (mean)	5.14	2.28	6.93	4.27	-1.27	-0.25

Table A9.8. Spitalfields Collection, inaccuracy, standard deviation and bias for each age group, males only
*not a mean, based on one individual only.

A.9.3 Coimbra Collection

The percentage of correct age estimates using the overall method for Coimbra males followed the parabolic pattern, with an initially fairly low percentage that rose until ages 50 to 59, and then decreasing with age. The females and the sexes pooled did not follow this pattern, instead beginning with a higher percentage at the 20 to 29 age group, dropping for the 30 to 39 year olds, increasing again, and then dropping steadily until the 80 to 89 age group, where the correct estimates increased again. Thus, the male and female percentages of correctly estimated ages did not correspond by age group, but the total percentages of correct age estimates were fairly close – 46.6% of females and 52.2% of males were aged correctly. Details are in Table A9.9, below.

Known Age Group	Number of Individuals			Correct Age Estimates (n)			Correct Age Estimates (%)		
	F	M	P	F	M	P	F	M	P
20-29	10	9	19	7	5	12	70.0	55.6	63.2
30-39	10	10	20	5	7	12	50.0	70.0	60.0
40-49	10	10	20	8	9	17	80.0	90.0	85.0
50-59	10	10	20	7	9	16	70.0	90.0	80.0
60-69	10	11	21	5	4	9	50.0	36.4	42.9
70-79	10	10	20	0	1	1	0.0	10.0	5.0
80-89	10	6	16	2	0	2	20.0	0.0	12.5
90-99	3	1	4	0	0	0	0.0	0.0	0.0
100+	--	--	--	--	--	--	--	--	--
Total	73	67	140	34	35	69	46.6	52.2	49.3

Table A9.9. Coimbra Collection, numbers and percentages of correct overall estimates
F: female; M: male; P: pooled (sexes); --: n/a.

The percentages of correct subjective age estimates did not follow the parabolic pattern. Instead, the pooled estimates generally decrease in percentage across the age range. There were some minor fluctuations for the 30 to 39 and 40 to 49 year olds and the 70 to 79 and 80 to 89 year olds. The male percentages of correct age estimation decreased across the age range, with no fluctuations, but with the same number and percentage of correctly aged 30 to 39 and 40 to 49 year olds. The females generally decreased in percentages of correct age estimation, but with fluctuations in the same age groups as noted for the sexes pooled (indeed, the females are the cause of the fluctuations for the sexes pooled). The total percentage of correct age estimates was higher for females, at 72.6 %, compared to males, at 61.2%, for a correct pooled total of 67.1% for the subjective age estimates.

Known Age Group	Number of Individuals			Correct Age Estimates (n)			Correct Age Estimates (%)		
	F	M	P	F	M	P	F	M	P
20-29	10	9	19	10	8	18	100.0	88.9	94.7
30-39	10	10	20	8	7	15	80.0	70.0	75.0
40-49	10	10	20	9	7	16	90.0	70.0	80.0
50-59	10	10	20	7	6	13	70.0	60.0	65.0
60-69	10	11	21	8	6	14	80.0	54.6	66.7
70-79	10	10	20	4	5	9	40.0	50.0	45.0
80-89	10	6	16	6	2	8	60.0	33.3	50.0
90-99	3	1	4	1	0	1	33.3	0.0	25.0
100+	--	--	--	--	--	--	--	--	--
Total	73	67	140	53	41	94	72.6	61.2	67.1

Table A9.10. Coimbra Collection, numbers and percentages of correct subjective estimates
F: female; M: male; P: pooled (sexes); --: n/a.

The overall inaccuracy values for females only (Table A9.11) did not follow the same pattern found for Spitalfields and the Coimbra sexes pooled; rather, the values fluctuated until the 50 to 59 age group, after which they increased steadily with age. Standard deviation did the same, but also dropped at the oldest age group. Bias also fluctuated somewhat until the 50 to 59 age group, where it again increased consistently with age. Overall bias was positive for the 20 to 29, 30 to 39, and 50 to 59 age groups, and negative for the 40 to 49 and older age groups. The subjective inaccuracy also fluctuated until the 60 to 69 age group; it then increased steadily, but the values were much smaller than for the overall estimates (for example, the 90 to 99 group had an overall inaccuracy of 18.67 compared to a subjective inaccuracy of 5.00). The subjective pattern for standard deviation was the same. Subjective bias also fluctuated until the 60 to 69 age group; from there on, bias steadily increased in value. Bias was positive for the 30 to 39 age group, and negative for the 40 to 49, and 70 to 79 and older age groups. Bias was neutral, at 0.00, for the 20 to 29, 50 to 59 and 60 to 69 age groups. Total mean inaccuracy, standard deviation and bias were all lower for the subjective estimates than for the overall estimates.

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	0.50	0.00	0.85	0.00	0.50	0.00
30-39	1.80	2.10	3.08	4.72	1.80	2.10
40-49	0.40	0.10	0.97	0.32	-0.20	-0.10
50-59	0.60	1.00	1.07	1.89	0.40	0.00
60-69	1.70	1.00	2.79	2.11	-1.70	0.00
70-79	9.80	4.40	6.23	5.15	-9.80	-4.00
80-89	13.30	3.70	8.71	5.33	-13.30	-3.70
90-99	18.67	5.00	5.51	4.58	-18.67	-5.00
100+	--	--	--	--	--	--
Total (mean)	4.62	1.89	7.03	1.89	-3.82	-0.99

Table A9.11. Coimbra Collection, inaccuracy, standard deviation and bias for each age group, females only

The males followed a different pattern for inaccuracy, standard deviation and bias, more similar to that seen in Spitalfields than for the Coimbra females (see Table A9.12). Overall inaccuracy and standard deviation both started fairly high at the 20 to 29 age group, and decreased until the 50 to 59 group, increasing again from the 60 to 69 age group onwards. The exceptions to this were the inaccuracy values for 90 to 99 year olds, as there was only one individual in the group, and standard deviation for the 80 to 89 age group, where there was a slight drop in value. Overall bias followed the same pattern, decreasing until the 50 to 59 age group, before steadily increasing from the 60 to 69 age group and older; the 90 to 99 age group did not break with the pattern of increasing value, but was the result of one individual only, so was not a group mean. Bias was positive from the 20 to 29 age group until the 40 to 49 age group, changing to negative at the 50 to 59 age group, and remaining negative for the rest of the age groups. Subjective inaccuracy, meanwhile, fluctuated across the entire age range, with a low of 0.33 for the 20 to 29 age group, to a high of 4.40 for the 70 to 79 age group. Standard deviation similarly fluctuated across the entire age range. Subjective bias for the males fluctuates for the first four age groups, and then steadily increased after the 60 to 69 age group. As with the overall bias, subjective bias was positive for the first three age groups, and then switched to negative from the 50 to 59 age group onwards. Overall total mean inaccuracy, standard deviation and bias were again higher than for the subjective estimates. Inaccuracy, standard deviation and bias were lower for the subjective estimates than for the overall estimates. The subjective female total mean values were all slightly lower than those of the males only. The overall total inaccuracy and standard deviation for females were also slightly lower than that of the males, while overall total female bias was slightly higher than that of the males.

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	4.78	0.33	7.63	1.00	4.78	0.33
30-39	2.30	1.60	3.77	3.24	2.30	1.20
40-49	0.60	1.00	1.90	2.00	0.60	0.80
50-59	0.50	4.00	1.58	7.09	-0.50	-1.80
60-69	3.36	2.82	3.35	4.56	-3.36	-0.82
70-79	11.10	4.40	6.76	6.60	-11.10	-3.60
80-89	20.67	4.33	5.5	6.89	-20.67	-4.33
90-99	24*	3*	--	--	-24*	-3*
100+	--	--	--	--	--	--
Total (mean)	5.57	2.58	7.79	2.58	-3.42	-1.03

Table A9.12. Coimbra Collection, inaccuracy, standard deviation and bias for each age group, males only
*not a mean, based on one individual only.

A.9.4 Lisbon Collection

The Lisbon Collection pattern of correct overall age estimates for the sexes pooled again rises over the first three age groups, then begins to decrease over the rest of the age range, starting from the 50 to 59 age group. No individuals from the oldest age groups were correctly aged (from the 70 to 79 age group and older). The female and male patterns of correct overall estimates were the same as that of the sexes pooled although, in total, more males than females were aged correctly (51.4% compared to 44.7%). Full details are in Table A9.13, below. The percentages of correct subjective age estimates did not follow the same pattern (see Table A9.14). For the sexes pooled, the percentages decrease over the first three age groups before rising briefly at 50 to 59, then fluctuating somewhat over the remaining age groups, with lower proportions of correct estimates than those of the younger age groups. The male and female proportions of correct age estimates fluctuate across the age range; the highest proportion of correct female estimates was 100.0%, for the 20 to 29 year olds, while that for the male was 100.0%, for the 30 to 39 year olds. The subjective estimates were more successful at correct age determination than were the overall estimates, for the sexes pooled (70.6% compared to 48.0%), for females alone (65.8% compared to 44.7%) and for males alone (75.7% compared to 51.4%). While none of the individuals aged 70 and over were aged correctly using overall estimates, 65.0% of the 70 to 79 year olds, and 55.0% of the 80 to 89 year olds were aged successfully using subjective estimates, as well as 14.3% of 90 to 99 year olds (all of whom were female).

Known Age Group	Number of Individuals			Correct Age Estimates (n)			Correct Age Estimates (%)		
	F	M	P	F	M	P	F	M	P
20-29	10	10	20	10	6	16	100.0	60.0	80.0
30-39	9	10	19	8	8	16	88.9	80.0	84.2
40-49	10	10	20	7	10	17	70.0	100.0	85.0
50-59	10	10	20	7	9	16	70.0	90.0	80.0
60-69	10	10	20	2	3	5	20.0	30.0	25.0
70-79	10	10	20	0	0	0	0.0	0.0	0.0
80-89	10	10	20	0	0	0	0.0	0.0	0.0
90-99	7	--	--	0	--	--	0.0	--	--
100+	--	--	--	--	--	--	--	--	--
Total	76	70	146	34	36	70	44.7	51.4	48.0

Table A9.13. Lisbon Collection, numbers and percentages of correct overall estimates
F: female; M: male; P: pooled (sexes); --: n/a.

Known Age Group	Number of Individuals			Correct Age Estimates (n)			Correct Age Estimates (%)		
	F	M	P	F	M	P	F	M	P
20-29	10	10	20	10	9	19	100.0	90.0	95.0
30-39	9	10	19	8	10	18	88.9	100.0	94.7
40-49	10	10	20	5	9	14	50.0	90.0	70.0
50-59	10	10	20	8	9	17	80.0	90.0	85.0
60-69	10	10	20	5	5	10	50.0	50.0	50.0
70-79	10	10	20	6	7	13	60.0	70.0	65.0
80-89	10	10	20	7	4	11	70.0	40.0	55.0
90-99	7	--	--	1	--	--	14.3	--	--
100+	--	--	--	--	--	--	--	--	--
Total	76	70	146	50	53	103	65.8	75.7	70.6

Table A9.14. Lisbon Collection, numbers and percentages of correct subjective estimates
F: female; M: male; P: pooled (sexes); --: n/a.

The sexes followed different patterns of overall inaccuracy (Tables A9.15 and A9.16 have details). For females, overall inaccuracy steadily rose over the age range, with a slight drop at the 50 to 59 age group before rising again to a peak at the 90 to 99 age group. Males, however, decreased in overall inaccuracy over the first three age groups, before rising again at the 50 to 59 age group and increasing. The female subjective inaccuracy fluctuated over the age range; the lowest value was 0.00, for the 20 to 29 year olds, while the highest was 8.30 for the 40 to 49 year olds. Male subjective inaccuracy was similar to the male overall inaccuracy pattern, decreasing from the 20 to 29 age group to the 30 to 39 year olds, rising again at the 40 to 49 age group, and increasing steadily across the remaining age groups. The female standard deviations for both overall and subjective estimates fluctuated over the age range. The male standard deviations for both overall and subjective estimates fluctuated over the younger age groups, but began to rise steadily from the 60 to 69 through to the 80 to 89 age groups. Female overall bias increased over the entire age range; the first three age groups had a positive bias, which changed to negative at the 50 to 59 age group. Subjective female bias also showed a weaker trend of increase across the age range, with the exception of the 40 to 49 age group. Bias was again positive for the first three age groups, and negative from the 50 to 59 age group onwards.

Total mean inaccuracy, standard deviation and bias were lower for the subjective female estimates than for the overall female estimates. Overall male bias had a different pattern, conforming instead to the parabolic pattern. Again, however, the bias for the first three age groups was positive, before changing to negative for the 50 to 59 age group and onwards. The male subjective bias pattern was parabolic. Here, however, only the first two age groups had a positive bias, changing to negative at the 40 to 49 age group and increasing consistently in value. Total mean male inaccuracy, standard deviation and bias were again lower for the subjective estimates compared to the overall. Female subjective inaccuracy and standard deviation and female overall inaccuracy and bias were slightly higher than the same for males, while female subjective bias and overall standard deviation were lower than those of the males.

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	0.00	0.00	0.00	0.00	0.00	0.00
30-39	0.89	0.22	2.67	0.67	0.89	0.22
40-49	2.30	8.30	4.52	14.43	1.50	5.70
50-59	1.50	0.90	2.95	2.02	-1.50	-0.30
60-69	4.90	3.00	4.92	5.08	-4.90	-1.80
70-79	10.90	1.80	5.20	3.74	-10.90	-1.20
80-89	18.20	1.40	4.57	2.67	-18.20	-1.40
90-99	23.00	2.86	5.69	2.48	-23.00	-2.86
100+	--	--	--	--	--	--
Total (mean)	7.20	2.32	8.75	6.19	-6.49	-0.10

Table A9.15. Lisbon Collection, inaccuracy, standard deviation and bias for each age group, females only

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	1.20	0.30	1.75	0.95	1.20	0.30
30-39	0.80	0.00	1.75	0.00	0.80	0.00
40-49	0.00	0.40	0.00	1.26	0.00	-0.40
50-59	1.00	0.90	3.16	2.85	-1.00	-0.90
60-69	6.20	1.80	5.47	2.35	-6.20	-1.00
70-79	15.10	3.30	3.54	5.87	-15.10	-3.30
80-89	21.50	7.40	7.95	10.18	-21.50	-7.40
90-99	--	--	--	--	--	--
100+	--	--	--	--	--	--
Total (mean)	6.54	2.01	8.86	5.11	-5.97	-1.81

Table A9.16. Lisbon Collection, inaccuracy, standard deviation and bias for each age group, males only

A.9.5 Dart Collection

The percentages of overall correct age estimates for the Dart Collection, for the sexes pooled, increased from the first to third age groups, before decreasing from the 50 to 59 age group onwards. The female and male patterns of correct overall age estimates followed the same pattern, although the decrease in percentage of successfully aged males began at the 60 to 69 age group instead of the 50 to 59 age group. The total percentage of correct age estimates for the sexes pooled was low, at 31.4%; the same for males alone (36.6%) is higher than that for females alone (26.0%). Table A9.17 has full details. The correct subjective estimates for the sexes pooled were fairly similar in proportions across the age range, only dropping steadily from the 70 to 79 age group onwards. For males alone, the subjective estimates were fairly similar across the younger to middle age groups, with the exception of a low of 40.0% for the 30 to 39 year olds, before dropping again from the 60 to 69 age group onwards; the females, meanwhile, had fairly similar percentages of correct age estimations until a decrease for the 90 to 99 year olds. Females were more often aged correctly using the subjective estimates than the males; the total percentages of correct age estimations for males, females and the sexes pooled were higher for the subjective age estimates than for the overall age estimates. Details for the subjective age estimates are in Table A9.18, below.

Known Age Group	Number of Individuals			Correct Age Estimates (n)			Correct Age Estimates (%)		
	F	M	P	F	M	P	F	M	P
20-29	10	10	20	2	5	7	20.0	50.0	35.0
30-39	10	10	20	4	7	11	40.0	70.0	55.0
40-49	10	10	20	6	8	14	60.0	80.0	70.0
50-59	10	10	20	4	8	12	40.0	80.0	60.0
60-69	10	10	20	3	2	5	30.0	20.0	25.0
70-79	10	10	20	1	0	1	10.0	0.0	5.0
80-89	10	10	20	0	0	0	0.0	0.0	0.0
90-99	6	9	15	0	0	0	0.0	0.0	0.0
100+	1	3	4	0	0	0	0.0	0.0	0.0
Total	77	82	159	20	30	50	26.0	36.6	31.4

Table A9.17. Dart Collection, numbers and percentages of correct overall estimates
F: female; M: male; P: pooled (sexes); --: n/a.

Known Age Group	Number of Individuals			Correct Age Estimates (n)			Correct Age Estimates (%)		
	F	M	P	F	M	P	F	M	P
20-29	10	10	20	6	7	13	60.0	70.0	65.0
30-39	10	10	20	7	4	11	70.0	40.0	55.0
40-49	10	10	20	6	7	13	60.0	70.0	65.0
50-59	10	10	20	6	7	13	60.0	70.0	65.0
60-69	10	10	20	7	6	13	70.0	60.0	65.0
70-79	10	10	20	7	4	11	70.0	40.0	55.0
80-89	10	10	20	7	2	9	70.0	20.0	45.0
90-99	6	9	15	2	2	4	33.3	22.2	26.7
100+	1	3	4	0	0	0	0.0	0.0	0.0
Total	77	82	159	48	39	87	62.3	47.6	54.7

Table A9.18. Dart Collection, numbers and percentages of correct subjective estimates
F: female; M: male; P: pooled (sexes); --: n/a.

For both females and males only, overall inaccuracy decreased over the youngest age groups, before increasing steadily. For females, this increase began at the 40 to 49 age group, while for males, the increase began at the 50 to 59 age group (Tables A9.19 and A9.20 have full details). Male and female overall standard deviation both fluctuated over the age range. Overall bias followed the same pattern as inaccuracy for both males and females, decreasing initially, before beginning to increase from ages 50 to 59 onwards. Subjective inaccuracy for both males and females fluctuated over the age range, although the oldest age groups have the highest inaccuracies.

Subjective standard deviation fluctuated over the age range for males and females. Subjective female bias fluctuated over the age range, although the oldest age groups did have the highest bias. The male subjective bias fluctuated until the 60 to 69 age group, after which bias increased steadily. Female subjective bias was positive for the 20 to 29, and 40 to 49 age groups, and was neutral for the 50 to 59 age group; the 30 to 39 and all age groups from 60 to 69 onwards were negative in bias. Male subjective bias followed the nearly-standard pattern of positive bias for the youngest three age groups, before changing to negative bias at the 50 to 59 age group, and remaining negative for all other age groups. Subjective female inaccuracy, standard deviation and bias were lower than for the overall estimates. The same was true for male subjective inaccuracy, standard deviation and bias compared to the same values for the overall estimates. The total mean female values for overall and subjective inaccuracy, standard deviation and bias were lower than for males.

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	5.20	1.30	4.44	2.06	5.20	1.30
30-39	1.90	1.30	2.51	2.16	1.10	-1.30
40-49	2.00	4.50	3.16	9.03	0.40	1.30
50-59	2.40	1.80	2.99	3.16	-2.00	0.00
60-69	9.60	4.30	8.22	8.64	-9.60	-3.90
70-79	12.50	3.40	7.55	5.68	-12.50	-3.40
80-89	17.10	1.30	7.22	2.31	-17.10	-1.30
90-99	20.17	5.83	6.82	6.77	-20.17	-5.83
100+	43*	22*	--	--	-43*	-22*
Total (mean)	8.71	3.06	9.23	5.95	-6.61	-1.69

Table A9.19. Dart Collection, inaccuracy, standard deviation and bias for each age group, females only
*not a mean, based on one individual only.

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	2.90	2.20	4.20	3.61	2.90	2.20
30-39	1.80	2.40	3.36	2.67	1.80	1.60
40-49	1.40	3.90	3.78	8.81	1.40	3.90
50-59	1.60	2.80	3.50	7.18	-1.60	-1.80
60-69	5.00	1.50	3.20	2.55	-5.00	-0.50
70-79	13.30	3.90	6.09	4.91	-13.30	-3.90
80-89	20.50	5.20	4.30	4.02	-20.50	-5.20
90-99	30.00	8.78	10.63	11.83	-30.00	-8.78
100+	49.67	30.00	3.51	11.79	-49.67	-30.00
Total (mean)	10.78	4.73	13.32	8.24	-9.29	-2.51

Table A9.20. Dart Collection, inaccuracy, standard deviation and bias for each age group, males only

A.9.6 Pretoria Collection

For the Pretoria Collection, the percentage of correct overall estimates for the sexes pooled was fairly similar from the 20 to 29 through to the 50 to 59 age groups, decreasing thereafter. No individuals were correctly aged from 70 to 79 through to 90 to 99 years old. Male and female patterns were the same as for the sexes pooled, and had values for each age group very similar to each other. The male, female and sexes pooled total percentage of correct age estimates were very similar (39.2%, 40.5% and 39.9%, respectively). Table A9.21 has full details. The percentages of correct subjective estimates for the sexes pooled fluctuated until the oldest age groups, which were relatively lower than the other age groups. The male and female percentages of correct age estimates also varied by age group, but with lower values for the oldest age groups. The total percentage of correct subjective age estimates for the sexes pooled was 68.2%; the male percentage was 66.2%, lower than that for the females alone, at 70.3%. Table A9.22 has full details.

Known Age Group	Number of Individuals			Correct Age Estimates (n)			Correct Age Estimates (%)		
	F	M	P	F	M	P	F	M	P
20-29	9	10	19	5	6	11	55.6	60.0	57.9
30-39	11	10	21	8	8	16	72.7	80.0	76.2
40-49	10	10	20	7	6	13	70.0	60.0	65.0
50-59	11	10	21	7	6	13	63.6	60.0	61.9
60-69	11	10	21	3	3	6	27.3	30.0	28.6
70-79	10	10	20	0	0	0	0.0	0.0	0.0
80-89	10	10	20	0	0	0	0.0	0.0	0.0
90-99	2	4	6	0	0	0	0.0	0.0	0.0
100+	--	--	--	--	--	--	--	--	--
Total	74	74	148	30	29	59	40.5	39.2	39.9

Table A9.21. Pretoria Collection, numbers and percentages of correct overall estimates
F: female; M: male; P: pooled (sexes); --: n/a.

Known Age Group	Number of Individuals			Correct Age Estimates (n)			Correct Age Estimates (%)		
	F	M	P	F	M	P	F	M	P
20-29	9	10	19	5	10	15	55.6	100.0	79.0
30-39	11	10	21	9	8	17	81.8	80.0	81.0
40-49	10	10	20	6	10	16	60.0	100.0	80.0
50-59	11	10	21	11	4	15	100.0	40.0	71.4
60-69	11	10	21	6	5	11	54.6	50.0	52.4
70-79	10	10	20	8	7	15	80.0	70.0	75.0
80-89	10	10	20	6	3	9	60.0	30.0	45.0
90-99	2	4	6	1	2	3	50.0	50.0	50.0
100+	--	--	--	--	--	--	--	--	--
Total	74	74	148	52	49	101	70.3	66.2	68.2

Table A9.22. Pretoria Collection, numbers and percentages of correct subjective estimates
F: female; M: male; P: pooled (sexes); --: n/a.

For females only, the overall inaccuracy decreased initially, with a minor fluctuation at the 40 to 49 and 50 to 59 age groups, before steadily increasing from the 60 to 69 age group onwards. Overall standard deviation fluctuated across the age range, while overall female bias followed the same pattern as that of overall female inaccuracy, peaking at the 90 to 99 age group. The first three age groups had positive biases, while the remaining older age groups all had negative biases. Female subjective inaccuracy fluctuated across the age range, although the oldest two age groups had the highest inaccuracies. Female subjective standard deviation decreased initially until reaching neutrality at the 50 to 59 age group, then fluctuated over the remaining age groups; here again, the oldest age groups had the highest standard deviations. Female subjective bias also fluctuated over the age range, but was highest for the oldest age groups. Bias was positive for the 20 to 29 through to the 50 to 59 age groups, as well as slightly positive for the 70 to 79 year olds (0.10); bias was negative for the 60 to 69, 80 to 89 and 90 to 99 age groups. Total mean female subjective inaccuracy, standard deviation and bias were lower than the same for the overall estimates. Table A9.23 has the details for females only.

The Pretoria male overall inaccuracy decreased initially, until the 40 to 49 age group, before rising steadily from the 50 to 59 age group onwards. Overall standard deviation fluctuated across the age range, but was higher for the oldest three age groups compared to the younger age groups. Overall male bias also decreased initially until the 40 to 49 age group, increasing from the 50 to 59 group onwards, to a high of -25.75 for the 90 to 99 age group. The overall bias was positive for the first three age groups and negative thereafter. The subjective male inaccuracy fluctuated throughout the age range, although the oldest age group (90 to 99) had the highest inaccuracy. Subjective standard deviation also fluctuated over the age range, but had two high points: 10.70 for the 30 to 39 year olds and 8.35 for the 90 to 99 year olds. Subjective bias fluctuated over the younger age groups, before rising steadily from the 40 to 49 year olds to the 90 to 99 year olds. Subjective bias was positive for the 30 to 39 and 50 to 59 groups, neutral for the 20 to 29 and 40 to 49 groups, and negative from the 60 to 69 group onwards. Total mean subjective inaccuracy, standard deviation and bias were lower than the same for the overall estimates. The female values for overall and subjective inaccuracy, standard deviation and bias were all lower than the same for males.

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	4.22	2.22	6.72	3.70	4.22	2.22
30-39	1.55	1.18	2.81	3.13	0.64	1.18
40-49	1.60	2.20	4.06	3.01	1.60	2.20
50-59	1.36	0.00	2.73	0.00	-1.36	0.00
60-69	6.27	2.27	4.58	3.41	-6.27	-1.55
70-79	13.40	0.30	5.38	0.67	-13.40	0.10
80-89	16.20	3.90	3.99	7.32	-16.2	-3.90
90-99	26.00	4.50	1.41	6.36	-26.00	-4.50
100+	--	--	--	--	--	--
Total (mean)	6.80	1.77	7.67	3.78	-5.01	-0.12

Table A9.23. Pretoria Collection, inaccuracy, standard deviation and bias for each age group, females only

Known Age Group	Inaccuracy		Standard Deviation		Bias	
	Overall	Subjective	Overall	Subjective	Overall	Subjective
20-29	2.50	0.00	3.37	0.00	2.50	0.00
30-39	1.70	3.60	4.42	10.70	1.70	3.20
40-49	1.10	0.00	1.73	0.00	0.50	0.00
50-59	1.60	1.70	3.13	2.06	-1.60	0.10
60-69	5.10	2.90	4.93	3.38	-5.10	-0.70
70-79	12.10	1.90	6.40	4.01	-12.10	-1.90
80-89	22.40	3.60	10.95	3.72	-22.40	-3.60
90-99	25.75	4.50	5.50	8.35	-25.75	-4.50
100+	--	--	--	--	--	--
Total (mean)	7.68	2.09	10.01	4.98	-6.32	-0.64

Table A9.24. Pretoria Collection, inaccuracy, standard deviation and bias for each age group, males only

Appendix 10: Age Heaping Bar Charts for Grant, Coimbra, Lisbon and Spitalfields

No evidence of age heaping was found in Grant, Coimbra, Lisbon or Spitalfields (Figures A10.1 to A10.4). Age heaping was found in Dart and Pretoria; bar charts and discussion are in Chapter 5.

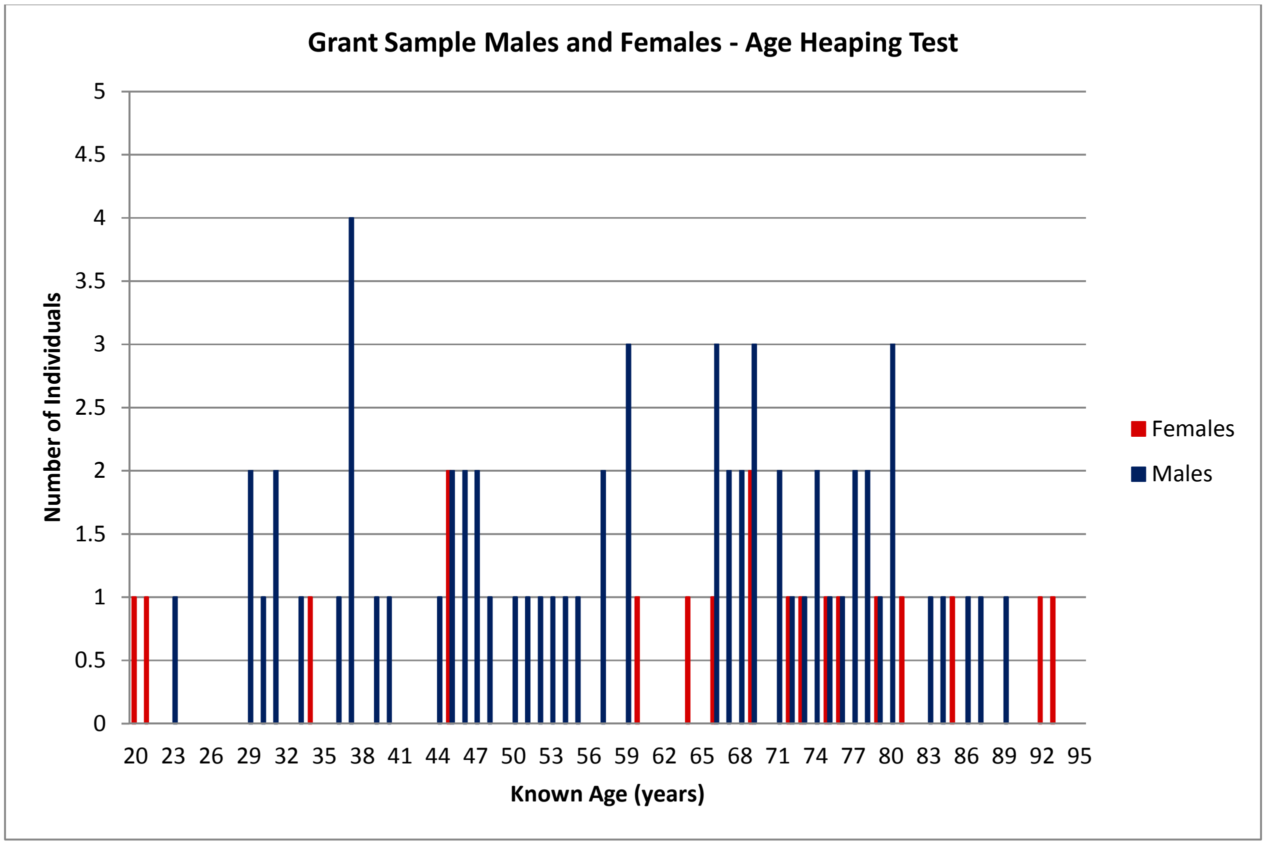


Figure A10.1. Documented Grant Collection ages-at-death, divided by sex

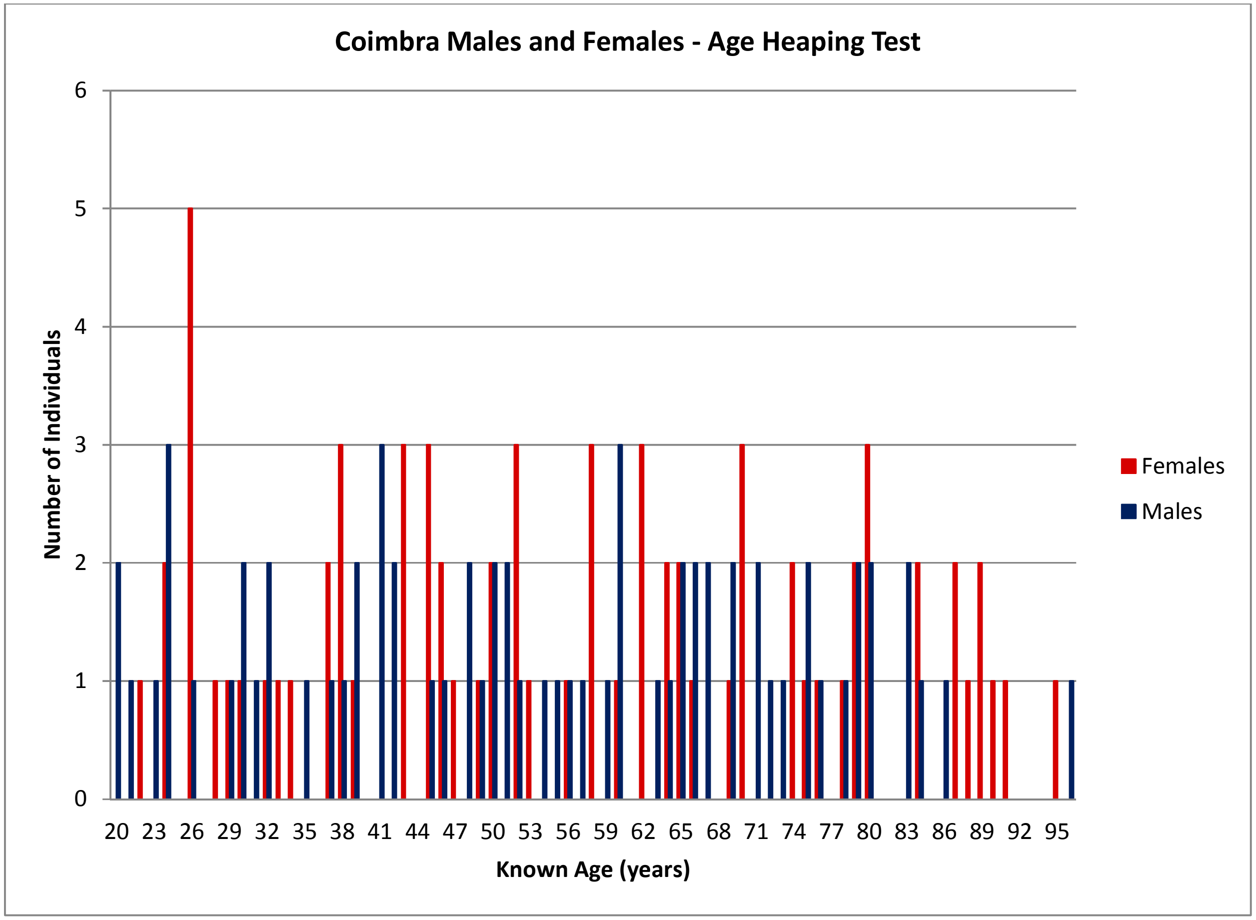


Figure A10.2. Documented Coimbra Collection ages-at-death, divided by sex

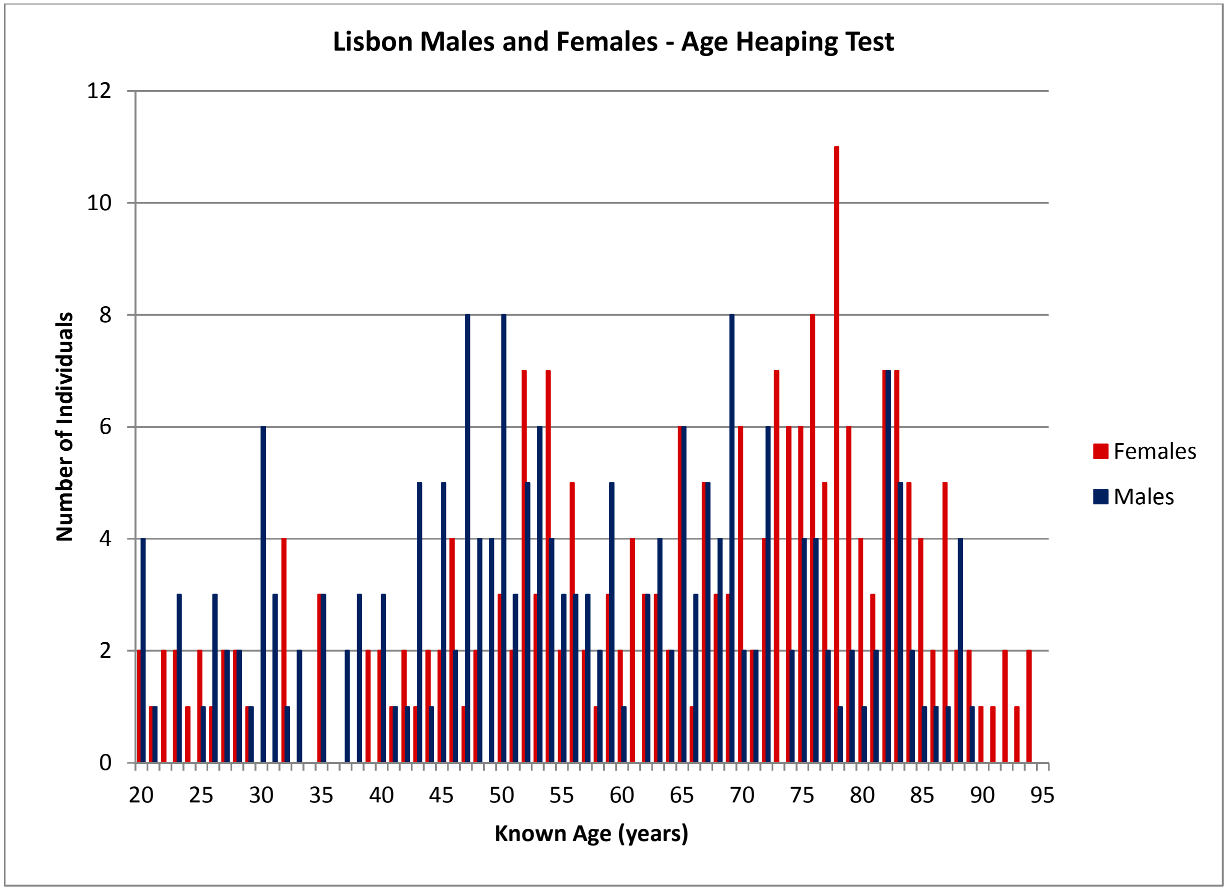


Figure A10.3. Documented Lisbon Collection ages-at-death, divided by sex

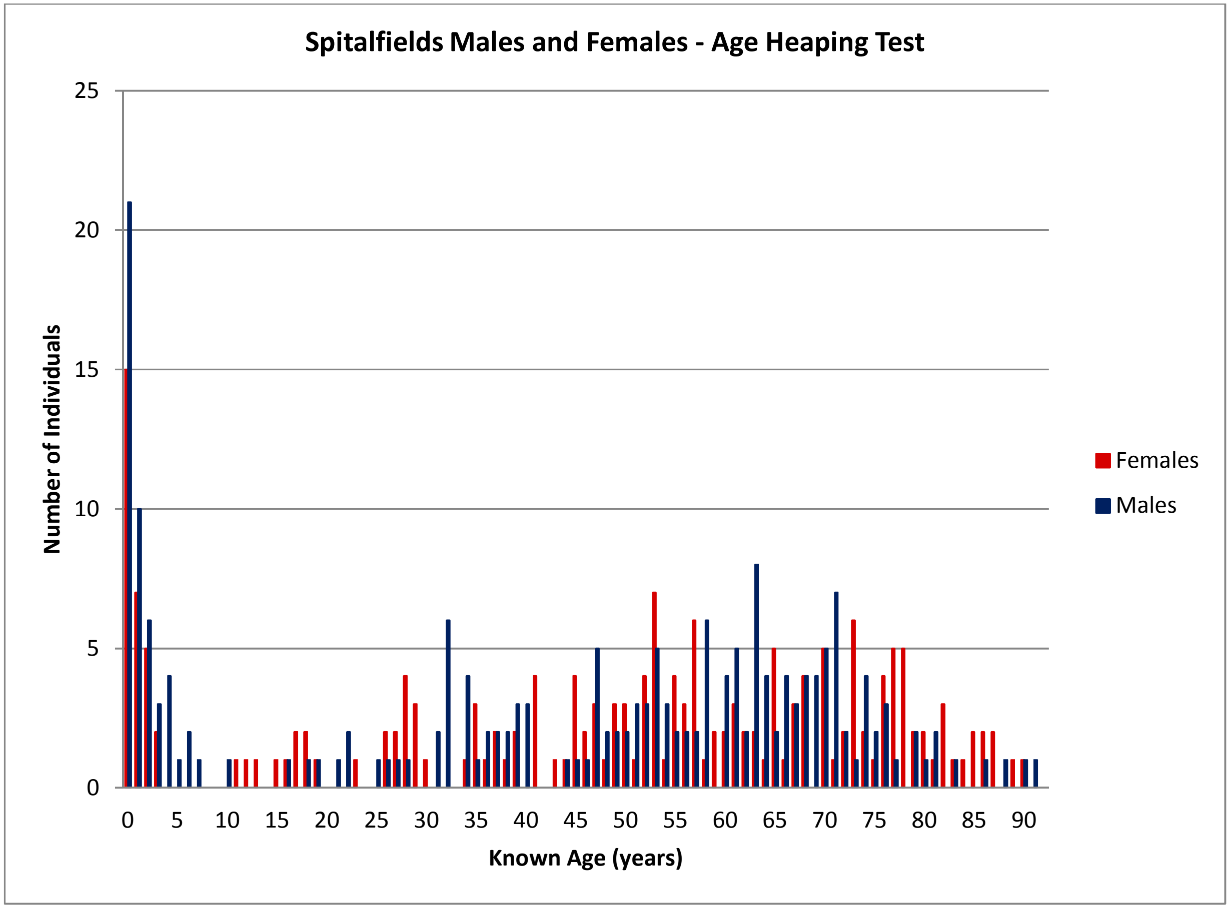


Figure A10.4. Documented Spitalfields Collection ages-at-death, divided by sex

Appendix 11: Mean Ages per Phase by Ageing Method for Each Collection (Sexes Pooled)

	Grant			Spitalfields			Lisbon			Coimbra			Dart			Pretoria		
Phase	*x*	S.D.	Range	*x*	S.D.	Range	*x*	S.D.	Range	*x*	S.D.	Range	*x*	S.D.	Range	*x*	S.D.	Range
1				28.50	7.94	22-40	21.75	2.36	20-25	22.17	2.40	20-26	25.67	5.57	20-32	23.00	1.00	22-24
2	29	8	21-37	38.00	11.18	23-49	24.90	3.03	21-31	24.00	2.45	20-26	30.00	3.95	24-35	25.86	3.89	21-32
3				40.70	13.65	29-70	34.56	9.81	23-56	33.67	12.87	23-70	36.90	15.84	22-65	39.86	16.45	24-85
4	48.10	19.48	23-86	50.50	15.27	27-85	51.14	16.64	25-93	47.89	12.39	26-87	51.51	21.49	26-108	50.04	17.25	25-96
5	65.51	13.71	39-89	61.04	13.37	35-86	68.41	14.87	27-90	62.80	16.35	29-96	68.56	17.40	38-104	62.51	17.53	28-90
6	69.65	13.26	44-93	74.45	10.88	44-91	78.53	14.28	46-94	77.67	10.77	58-95	84.31	11.99	57-96	78.14	11.34	57-94

Table A11.1. Mean ages, standard deviation and age range for Suchey-Brooks pubic symphysis phases for each collection

x: mean age; S.D.: standard deviation; range: whole age range

	Grant			Spitalfields			Lisbon			Coimbra			Dart			Pretoria		
Phase	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range
20-24							23.78	3.80	20-32	21.00	1.41	20-22	28.00	6.57	20-36	24.29	3.25	21-31
25-29				29.33	6.87	22-41	31.07	9.27	20-55	29.70	8.48	20-48	35.92	15.28	20-66	31.77	8.27	22-51
30-34				39.87	14.21	23-70	42.53	13.04	25-82	40.95	13.09	24-83	47.72	20.76	22-105	44.60	15.99	24-85
35-39				44.83	11.05	28-67	61.06	16.89	29-87	55.63	15.27	30-89	62.55	21.02	28-108	62.58	17.29	25-94
40-44	56.31	18.92	29-81	55.29	14.89	29-79	68.55	14.20	43-94	60.97	15.39	35-88	67.50	18.64	26-94	66.94	15.44	37-89
45-49	64.32	17.04	23-87	64.61	15.85	32-91	74.63	11.89	53-91	70.43	14.70	39-96	65.33	22.75	35-88	68.22	15.63	35-91
50-60	65.80	12.64	44-84	67.28	13.75	43-88	72.64	16.86	46-92	77.62	10.50	54-96	66.70	23.44	35-93	77.36	12.77	54-96
60+	77.22	12.31	54-93	70.17	15.75	53-87							89.00	8.40	75-96			

Table A11.2. Mean ages, standard deviation and age range for Meindl-Lovejoy auricular surface phases for each collection

x: mean age; S.D.: standard deviation; range: whole age range

	Grant			Spitalfields			Lisbon			Coimbra			Dart			Pretoria		
Phase	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range
I							20.50	0.71	20-21									
II				31.13	6.79	22-41	26.25	5.43	20-42	25.63	8.05	20-45	31.29	11.12	20-66	26.55	4.84	21-38
III				35.00	9.61	23-53	38.00	8.99	27-65	33.28	7.65	24-48	42.21	17.08	20-75	43.15	15.04	24-86
IV				43.24	13.77	26-70	52.27	17.58	25-87	46.60	17.02	24-89	58.34	21.91	25-105	54.65	16.30	25-90
V	52.40	14.11	34-81	49.11	13.20	28-71	67.70	11.79	42-90	56.92	15.40	26-88	62.66	20.99	22-108	65.59	17.32	28-94
VI	64.75	16.95	29-86	65.70	13.74	32-91	71.83	14.88	43-94	67.38	13.28	45-95	72.52	18.50	35-95	72.56	14.90	40-96
VII	69.21	20.11	23-93	69.55	14.36	43-88	76.09	16.38	46-94	77.80	13.24	48-96	81.33	16.51	53-96	83.00	9.80	69-96

Table A11.3. Mean ages, standard deviation and age range for Buckberry-Chamberlain auricular surface phases for each collection

x: mean age; S.D.: standard deviation; range: whole age range

	Grant			Spitalfields			Lisbon			Coimbra			Dart			Pretoria		
Score	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range
1																		
2																		
3																		
4							43.80	25.63	21-83									
5							41.86	21.05	20-94	44.35	22.84	20-89	44.52	24.16	20-86	44.83	20.27	26-89
6	64.00	19.60	23-93	48.45	21.80	25-89	59.68	21.41	23-93	49.09	21.06	23-89	58.49	23.58	25-108	50.31	21.00	21-88
7	63.03	15.82	33-89	60.16	17.61	21-91	62.38	18.65	29-94	62.06	16.05	26-95	65.09	20.50	28-104	62.78	18.77	22-96
8																		

Table A11.4. Mean ages, standard deviation and age range for lateral-anterior cranial suture closure phases for each collection

x: mean age; S.D.: standard deviation; range: whole age range

	Grant			Spitalfields			Lisbon			Coimbra			Dart			Pretoria		
Score	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range	x	S.D.	Range
1																		
2													28.00	5.51	20-36			
3				42.25	24.61	26-78	31.90	20.70	20-86	34.67	14.51	20-69	49.00	24.72	20-86	38.44	21.56	23-86
4				44.06	15.69	25-85	52.81	23.45	21-94	50.22	20.28	20-90	56.57	20.02	20-91	52.89	20.47	22-94
5	72.33	7.57	67-81	57.97	20.20	21-91	61.81	16.88	29-89	57.60	19.63	24-95	62.86	25.18	27-108	60.13	21.24	21-96
6				62.73	15.58	34-87	60.93	18.34	33-94	62.46	15.56	26-86	67.46	20.29	29-96	63.20	15.85	35-91
7				68.08	15.58	48-86	67.82	23.43	31-92	68.44	16.18	45-96	73.55	16.98	33-100	59.50	13.74	40-76

Table A11.5. Mean ages, standard deviation and age range for vault cranial suture closure phases for each collection

x: mean age; S.D.: standard deviation; range: whole age range

	Grant			Spitalfields			Lisbon			Coimbra			Dart			Pretoria		
Score	*x*	S.D.	Range	*x*	S.D.	Range	*x*	S.D.	Range	*x*	S.D.	Range	*x*	S.D.	Range	*x*	S.D.	Range
3																		
4																		
5													30.00	1.63	28-32	27.00	4.24	24-30
6				43.20	19.62	25-80							29.67	2.08	29-32	29.44	8.09	22-47
7				44.33	16.84	22-68	33.50	8.83	22-46	45.25	18.93	29-70	48.40	25.95	24-87	36.14	16.93	24-72
8				44.70	15.39	29-68	34.50	9.75	21-48	51.45	17.16	28-88	49.40	18.34	33-75	45.73	11.26	28-65
9				61.17	18.28	34-87	46.00	13.29	27-65	51.17	20.11	32-90				41.67	17.22	21-72
10				70.29	14.28	50-86	65.89	18.07	43-93	53.50	24.60	29-89	70.00	12.83	49-84	62.50	18.02	43-90
11				62.00	18.63	27-89	57.20	12.52	47-79	67.44	13.22	63-86	73.50	25.38	31-105	59.60	16.07	38-80
12													68.50	17.71	52-91			
13																		
14																		
15																		

Table A11.6. Mean ages, standard deviation and age range for sternal end of the fourth rib phases for each collection

x: mean age; S.D.: standard deviation; range: whole age range

Appendix 12: Boxplots for Meindl-Lovejoy Auricular Surface Method by Phase for Each Collection

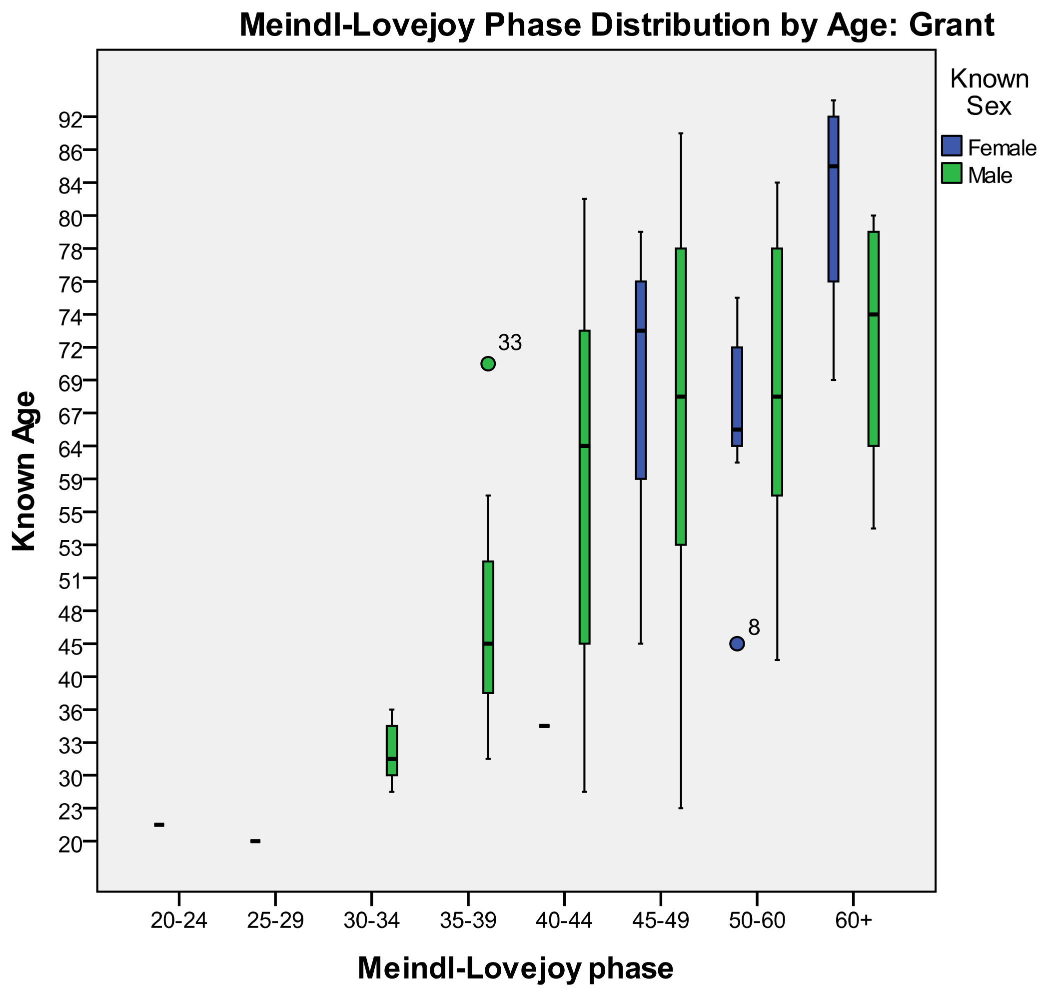


Figure A12.1. Boxplot of Meindl-Lovejoy auricular surface phase distribution by age for the Grant

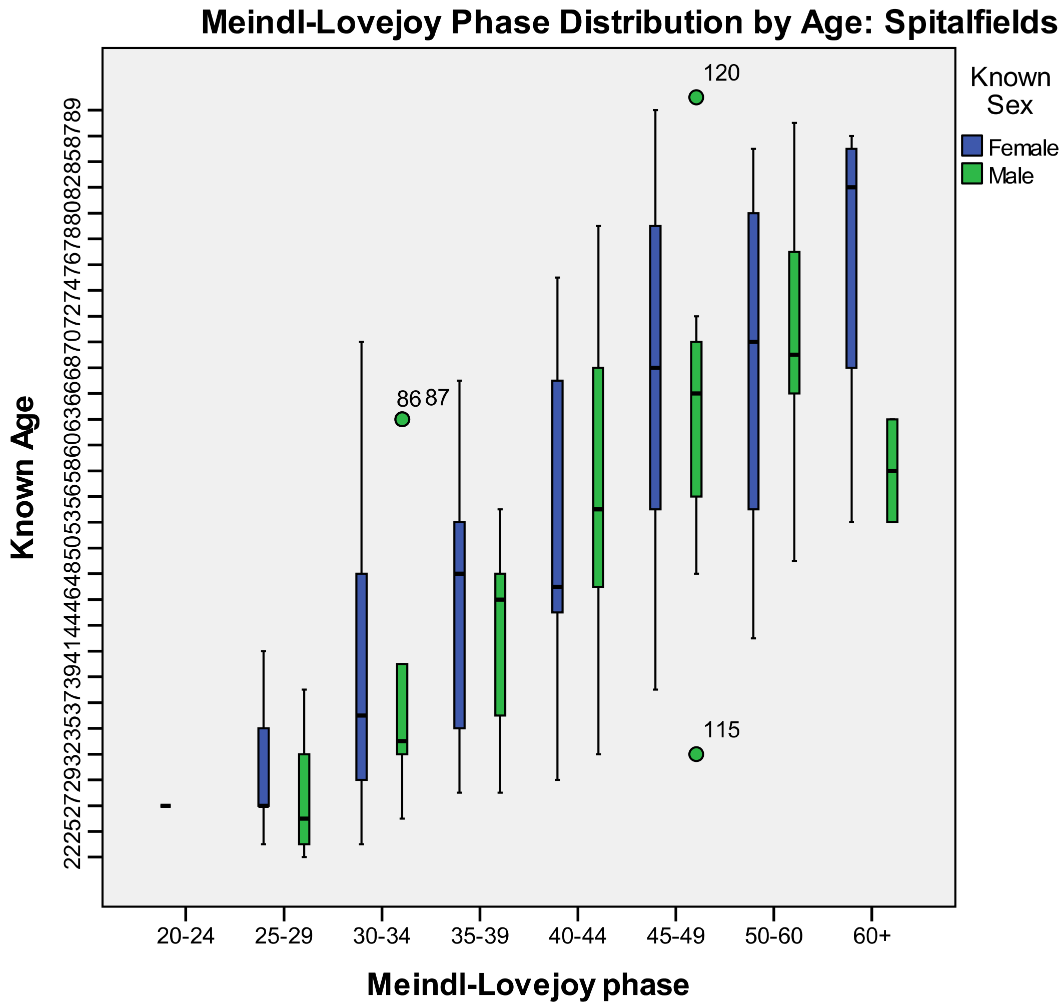


Figure A12.2. Boxplot of Meindl-Lovejoy auricular surface phase distribution by age for the Spitalfields

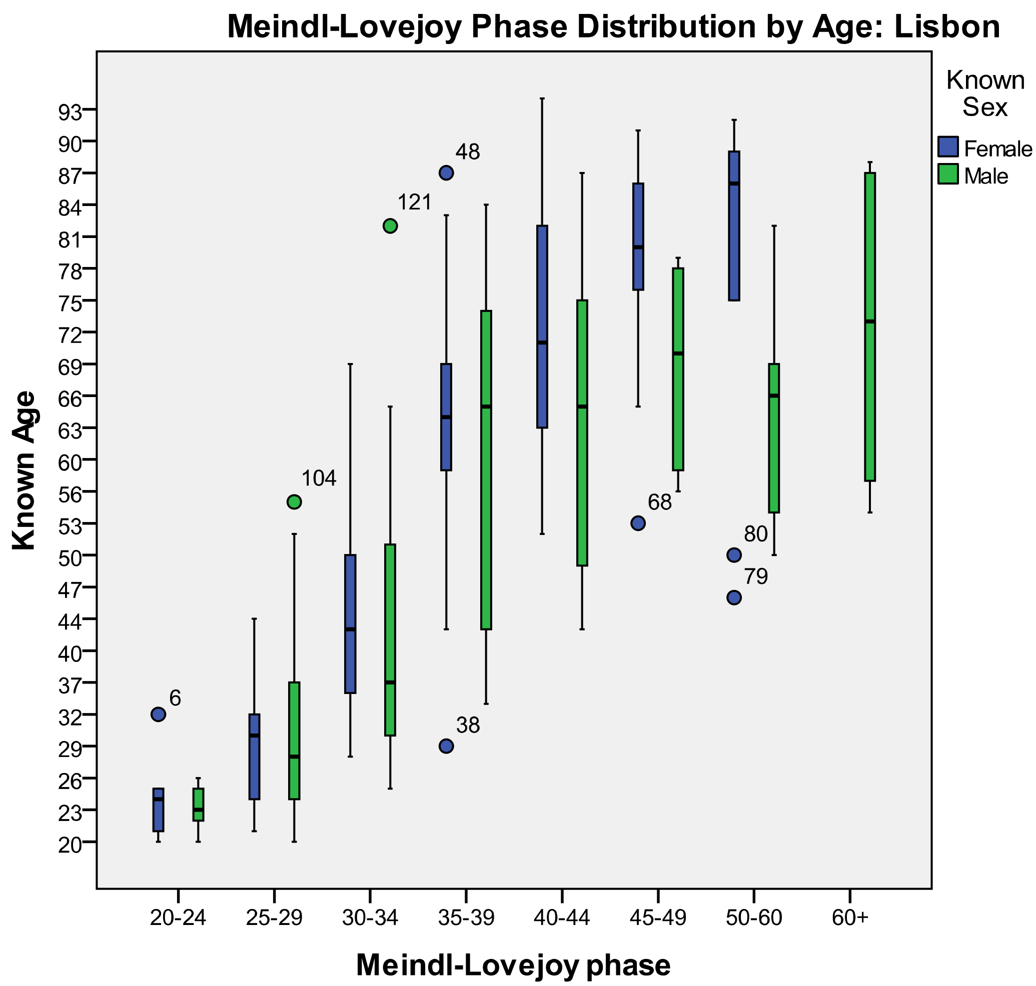


Figure A12.3. Boxplot of Meindl-Lovejoy auricular surface phase distribution by age for Lisbon

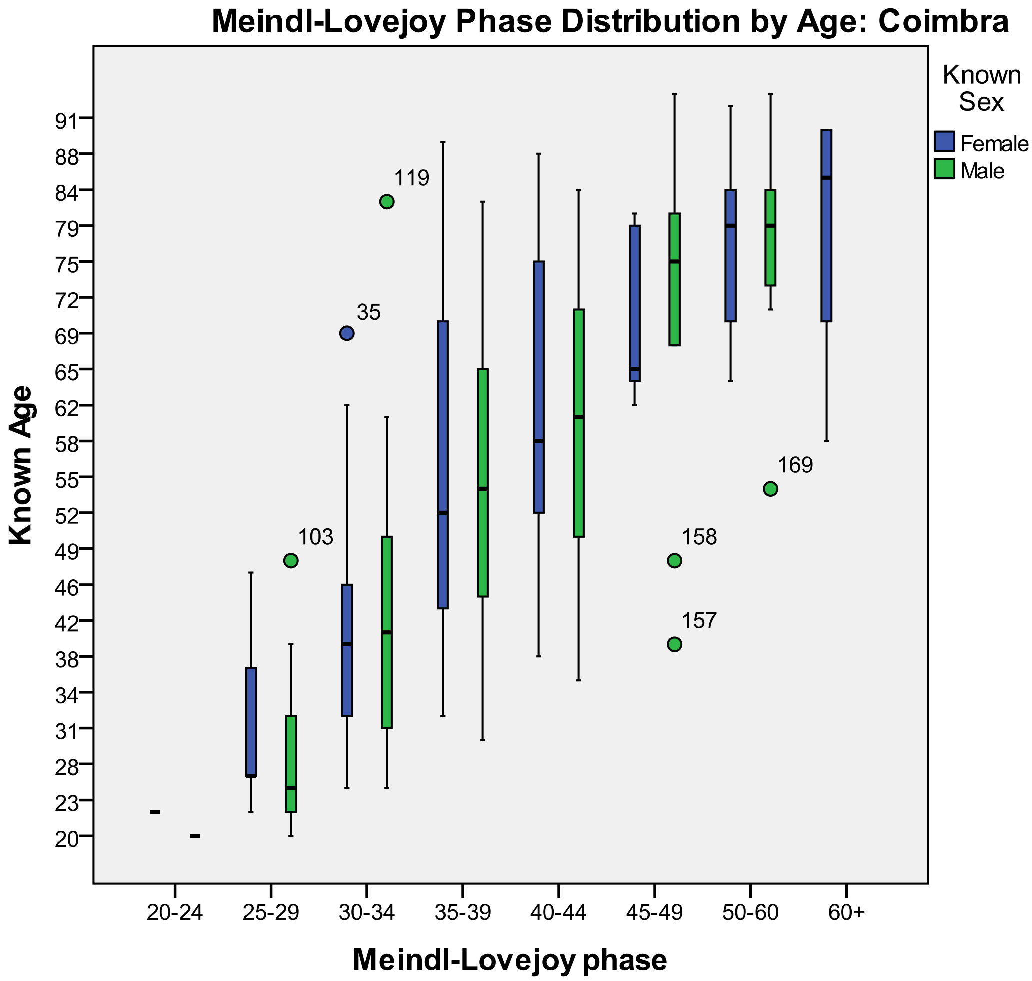


Figure A12.4. Boxplot of Meindl-Lovejoy auricular surface phase distribution by age for Coimbra

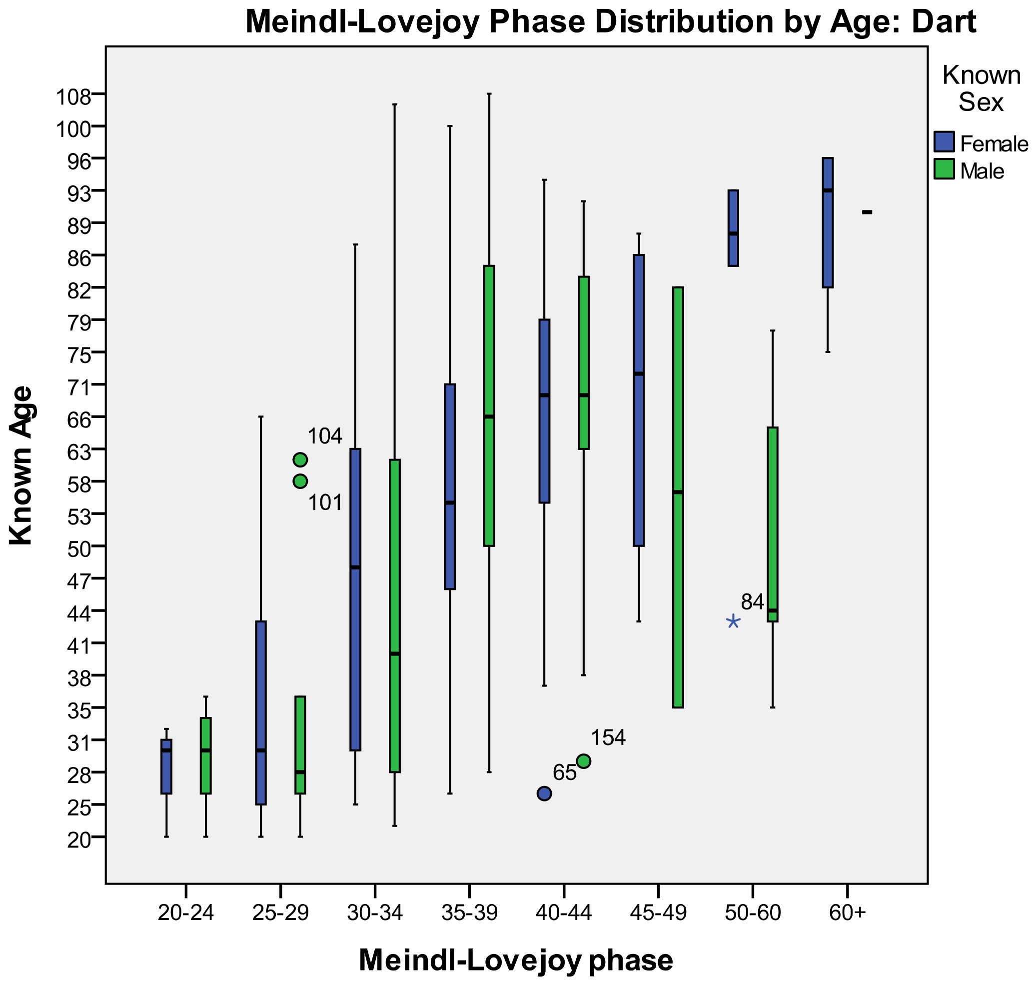


Figure A12.5. Boxplot of Meindl-Lovejoy auricular surface phase distribution by age for Dart

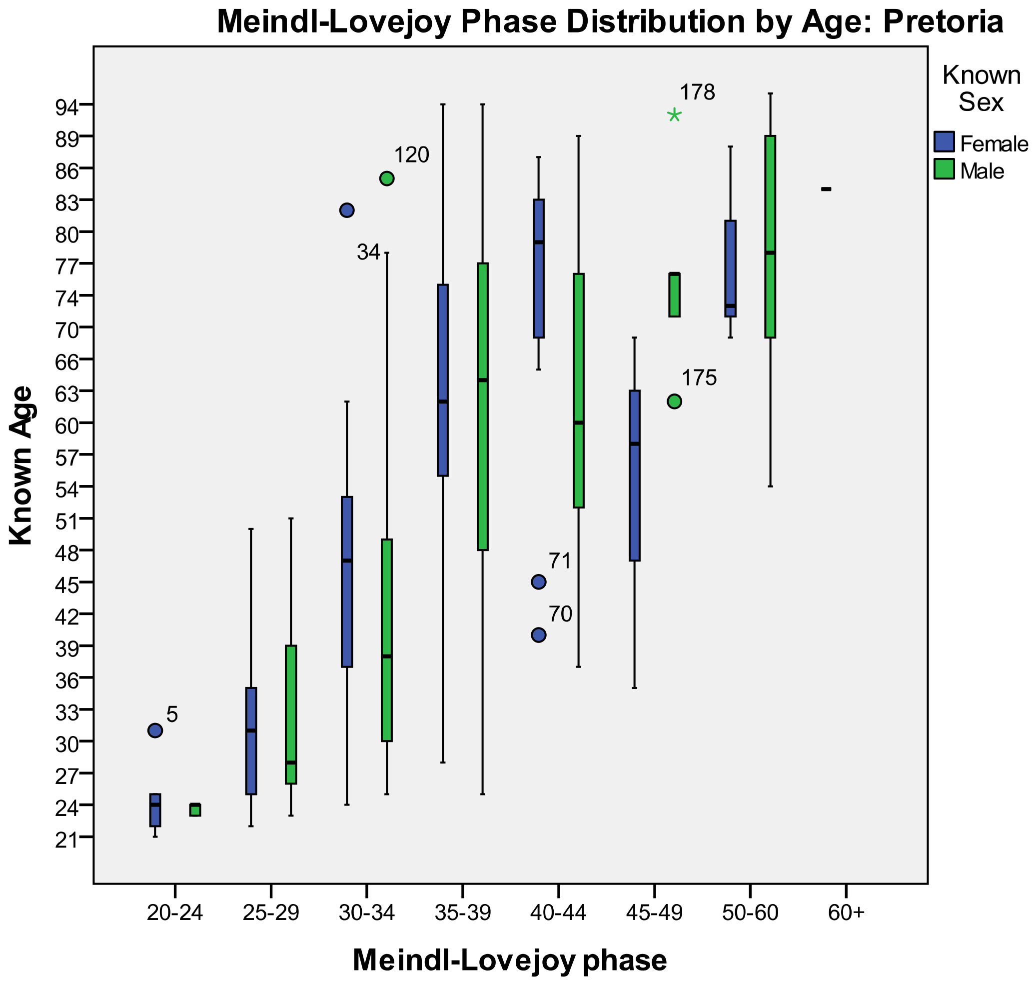


Figure A12.6. Boxplot of Meindl-Lovejoy auricular surface phase distribution by age group for Pretoria

Bibliography

Acsádi G, Nemeskéri J. 1970. *History of Human Life Span and Mortality*. Akadémiai Kiadó: Budapest.

Adams M, Reeve J. 1987. Excavations at Christ Church, Spitalfields 1984-6. *Antiquity* 61: 247-256.

Aiello LC, Molleson T. 1993. Are microscopic ageing techniques more accurate than macroscopic ageing techniques? *Journal of Archaeological Science* 20: 689-704.

Aktas EO, Koçak A, Aktas S, Yemisçigil A. 2004. Intercostal variation for age estimation – are the standards for the right 4th rib applicable for other ribs? *Collegium Antropologicum* 28: 267-272.

Albanese J. 2008. *Personal communication*.

Albanese J. 2003a. A metric method for sex determination using the hipbone and the femur. *Journal of Forensic Sciences* 48: 1-11.

Albanese J. 2003b. Identified Skeletal Reference Collections and the Study of Human Variation. *Unpublished PhD Thesis, McMaster University, Hamilton, Canada.*

Albanese J, Cardoso HFV, Saunders SR. 2005. Universal methodology for developing univariate sample-specific sex determination methods: an example using the epicondylar breadth of the humerus. *Journal of Archaeological Science* 32: 143-152.

Albanese J, Eklics G, Tuck A. 2008. A metric method for sex determination using the proximal femur and fragmentary hipbone. *Journal of Forensic Sciences* 53: 1283-1288.

Aleman I, Irurite J, Valencia AR, Martínez A, López-Lázaro S. 2012. Brief communication: the Granada Osteological Collection of Identified Infants and Young Children. *American Journal of Physical Anthropology* 149: 606-610.

Alesan A, Malgosa A, Simó C. 1999. Looking into the demography of an Iron Age population in the Western Mediterranean. I. Mortality. *American Journal of Physical Anthropology*. 110: 285-301.

Algee-Hewitt BFB, Weisensee KE, Milner GR. 2008. Age is subjective: a non-traditional method of age estimation for the adult skeleton. (Abstract) Paper presented at the 77th Annual Meeting of the American Association of Physical Anthropologists. *American Journal of Physical Anthropology* 135: 58.

Ali RS, MacLaughlin SM. 1991. Sex identification from the auricular surface of the adult human ilium. *International Journal of Osteoarchaeology* 1: 57-61.

Allen TC. 2012. Genetic susceptibility to lung cancer. *Molecular Pathology of Lung Cancer* 6: 7-25.

Anderson BE. 1990. Ventral arc of the os pubis: anatomical and developmental considerations. *American Journal of Physical Anthropology* 83: 449-458.

Anderson AS, Loeser RF. 2010. Why is osteoarthritis an age-related disease? *Best Practice & Research Clinical Rheumatology* 24: 15-26.

Armelagos GJ, Van Gerven DP. 2003. A century of skeletal biology and paleopathology: contrasts, contradictions, and conflicts. *American Anthropologist* 105: 53-64.

Asala SA. 2001. Sex determination from the head of the femur of South African whites and blacks. *Forensic Science International* 117: 15-22.

Asala SA. 2002. The efficiency of the demarking point of the femoral head as a sex determining parameter. *Forensic Science International* 127: 114-118.

Asala SA, Bidmos MA, Dayal MR. 2004. Discriminant function sexing of fragmentary femur of South African blacks. *Forensic Science International* 145: 25-29.

Ashley-Montagu MF. 1938. Aging of the skull. *American Journal of Physical Anthropology* 23: 355-375.

Audit Commission. 2008. Don't Stop Me Now – Preparing for an ageing population. Local government National Report, July 2008.

Aykroyd RG, Lucy D, Pollard AM, Roberts CA. 1999. Nasty, brutish, but not necessarily short: a reconsideration of the statistical methods used to calculate age-at-death from adult human skeletal and dental age indicators. *American Antiquity* 64: 55-70.

Baccino E, Schmitt A. 2006. Determination of adult age-at-death in the forensic context. In *Forensic Anthropology and Medicine: Complementary Sciences from Recovery to Cause of Death*, Schmitt A, Cunha E, Pinheiro J (eds.). Humana Press Inc.: New Jersey; 259-280.

Baccino E, Ubelaker DH, Hayek L-AC, Zerilli A. 1999. Evaluation of seven methods of estimating age-at-death from mature human skeletal remains. *Journal of Forensic Sciences*. 44: 931-936.

Bang G, Ramm E. 1970. Determination of age in humans from root dentine transparency. *Acta Odontologica Scandinavia* 28: 3-35.

Barrier ILO, L'Abbé EN. 2008. Sex determination from the radius and ulna in a modern South African sample. *Forensic Science International* 179: 85.e1-85.e7.

Bassuk SS, Berkman LF, Amick III BC. 2002. Socioeconomic status and mortality among the elderly: findings from four US communities. *American Journal of Epidemiology* 155: 520-533.

Beard J, Biggs S, Bloom D, Fried L, Hogan P, Kalache A, Olshansky J. 2012. Global population ageing: peril or promise? Program on the Global Demography of Aging Working Paper No. 89. Harvard Initiative for Global Health.

Bedford ME, Russell KF, Lovejoy CO, Meindl RS, Simpson SW, Stuart-Macadam PL. 1993. Test of multifactorial aging method using skeletons with known ages-at-death from the Grant Collection. *American Journal of Physical Anthropology* 91: 287-297.

Belcastro MG, Rastelli E, Mariotti V. 2008. Variation of the degree of sacral vertebral body fusion in adulthood in two European modern skeletal collections. *American Journal of Physical Anthropology* 135: 149-160.

Bell JT, Tsai P-C, Yang T-P, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A, Shin S-Y, Dempster EL, Murray RM, Grundberg E, Hedman AK, Nica A, Small KS, The MuTHER Consortium, Dermitzakis ET, McCarthy MI, Mill J, Spector TD and Deloukas P. 2012. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genetics* 8: e1002629. doi:10.1371/journal.pgen.1002629 (accessed 05/03/2013).

Bello SM, Thomann A, Signoli M, Dutour O, Andrews P. 2006. Age and sex bias in the reconstruction of past population structures. *American Journal of Physical Anthropology* 129: 24-38.

Berg GE. 2008. Pubic bone age estimation in adult women. *Journal of Forensic Sciences* 53: 569-577.

Beyer-Olsen EMS, Bang G, Sellevold BJ. 1994. Dental root dentine translucency used in age determination of Medieval Norwegians from Trondheim. *International Journal of Osteoarchaeology* 4: 305-310.

Bidmos M, Asala S. 2005. Calcaneal measurement in estimation of stature of South African blacks. *American Journal of Physical Anthropology* 126: 335-342.

Bielinski SJ, Pankow JS, Rasmussen-Torvik LJ, Bailey K, Li M, Selvin E, Couper D, Vazquez G, Brancati F. 2012. Strength of association for incident diabetes risk factors according to diabetes case definitions. *American Journal of Epidemiology* 175: 466-472.

Billings B. 2009. *Personal communication*.

Birkby WH. 1966. An evaluation of race and sex identification from cranial measurements. *American Journal of Physical Anthropology* 24: 21-28.

Blalock, HM. 1972. *Social Statistics*, 2nd Ed. McGraw-Hill Company: New York.

Bocquet-Appel JP, Masset C. 1982. Farewell to paleodemography. *Journal of Human Evolution* 11: 321-333.

Bogin B, Loucky J. 1997. Plasticity, political economy, and physical growth status of Guatemala Maya children living in the United States. *American Journal of Physical Anthropology* 102: 17-32.

Boldsen JL, Milner GR, Konigsberg LW, Wood JW. 2002. Transition analysis: a new method for estimating age from skeletons. In *Paleodemography – Age Distributions from Skeletal Samples*, Hoppa RD, Vaupel JW (eds.). Cambridge University Press: Cambridge; 73-106.

Borgognini Tarli SM, Repetto E. 1986. Methodological considerations on the study of sexual dimorphism in past human populations. *Human Evolution* 1: 51-66.

Borkan GA. 1986. Biological age assessment in adulthood. In *The Biology of Human Ageing*, Society for the Study of Human Biology Symposium 25, Bittles AH, Collins KJ (eds.). Cambridge University Press: Cambridge; 81-94.

Borkan GA, Hults DE, Mayer PJ. 1982. Physical anthropological approaches to aging. *Yearbook of Physical Anthropology* 25: 181-202.

Borkan GA, Norris AH. 1980. Biological age in adulthood: comparison of active and inactive US males. *Human Biology* 52: 787-802.

Bouvier M, Ubelaker DH. 1977. A comparison of two methods for the microscopic determination of age-at-death. *American Journal of Physical Anthropology* 46: 391-394.

Bowling A, Grundy E. 2009. Differentials in mortality up to 20 years after baseline interview among older people in East London and Essex. *Age and Ageing* 38: 51-55.

Bridges PS. 1989. Changes in activities with the shift to agriculture in the Southeastern United States. *Current Anthropology* 30: 385-394.

Broman GE, Trotter M, Peterson RR. 1958. The density of selected bones of the human skeleton. *American Journal of Physical Anthropology* 16: 197-211.

Brooks ST. 1955. Skeletal age-at-death: the reliability of cranial and pubic age indicators. *American Journal of Physical Anthropology* 13: 567-597.

Brooks ST, Suchey JM. 1990. Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks method. *Human Evolution* 5: 227-238.

Brothwell D. 1981. *Digging Up Bones*, 3rd edition. London: British Museum of Natural History.

Brown RA, Armelagos GJ. 2001. Apportionment of racial diversity: a review. *Evolutionary Anthropology: Issues, News and Reviews* 10: 34-40.

Brown P, Maeda T. 2004. Post-Pleistocene diachronic change in East Asian facial skeletons: the size, shape and volume of the orbits. *Anthropological Science* 112: 29-40.

Brunner C, Kissling R, Jacob HAC. 1991. The effects of morphology and histopathologic findings on the mobility of the sacroiliac joint. *Spine* 16: 1111-1117.

Bruzek J. 2002. A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology* 117: 157-168.

Bruzek J, Murail P. 2006. Methodology and reliability of sex determination from the skeleton. In *Forensic Anthropology and Medicine – Complementary Sciences from Recovery to Cause of Death,* Schmitt A, Cunha E, Pinheiro J (eds.). Humana Press, Inc.: New Jersey; 225-242.

Buckberry JL, Chamberlain AT. 2002. Age estimation from the auricular surface of the ilium: a revised method. *American Journal of Physical Anthropology* 119: 231-239.

Budinoff LC, Tague RG. 1990. Anatomical and developmental bases for the ventral arc of the human pubis. *American Journal of Physical Anthropology* 82: 73-79.

Buikstra JE, Gordon CC. 1981. The study and restudy of human skeletal series: the importance of long-term curation. *Annals of the New York Academy of Sciences* 376: 449-465.

Buikstra JE, Konigsberg LW. 1985. Paleodemography: critiques and controversies. *American Anthropologist* 87: 316-333.

Buikstra JE, Ubelaker DH. (eds.)1994. *Standards for Data Collection from Human Skeletal Remains.* Arkansas Archaeological Survey: Fayetteville.

Buretić-Tomijanović A, Ostojić S, Kapović M. 2006. Secular change of craniofacial measures in Croatian younger adults. *American Journal of Human Biology* 18: 668-675.

Bytheway JA, Ross AH. 2010. A geometric morphometric approach to sex determination of the human adult os coxa. *Journal of Forensic Sciences* 55: 859-864.

Caldwell WE, Moloy HC. 1932. Sexual variations in the pelvis. *Science* 76: 37-40.

Cameron N, Demerath EW. 2002. Critical periods in human growth and their relationship to diseases of aging. *Yearbook of Physical Anthropology* 45: 159-184.

Campanacho V, Santos AL, Cardoso HFV. 2012. Assessing the influence of occupational and physical activity on the rate of degenerative change of the pubic symphysis in Portuguese males from the 19th to 20th century. *American Journal of Physical Anthropology* 148: 371-378.

Cannet C, Baraybar JP, Kolopp M, Meyer P, Ludes B. 2011. Histomorphometric estimation of age in paraffin-embedded ribs: a feasibility study. *International Journal of Legal Medicine* 125: 493-502.

Cardoso FA. 2008. A Portrait of Gender in Two 19th and 20th Century Portuguese Populations: A Palaeopathological Perspective. Unpublished PhD thesis, Durham University.

Cardoso HFV. 2005. Patterns of Growth and Development of the Human Skeleton and Dentition in Relation to Environmental Quality: A Biocultural Analysis of a Sample of 20th Century Portuguese Subadult Documented Skeletons. *Unpublished PhD Thesis, McMaster University, Hamilton, Canada.*

Cardoso HFV. 2006. Brief communication: the collection of identified human skeletons housed at Bocage Museum (National Museum of Natural History), Lisbon, Portugal. *American Journal of Physical Anthropology* 129: 172-176.

Cardoso HFV. 2007a. Environmental effects on skeletal versus dental development: using a documented subadult skeletal sample to test a basic assumption in human osteological research. *American Journal of Physical Anthropology* 132: 223-233.

Cardoso HFV. 2007b. A test of the differential accuracy of the maxillary versus the mandibular dentition in age estimations of immature skeletal remains based on developing tooth length. *Journal of Forensic Sciences* 52: 434-437.

Cardoso HFV. 2008a. Epiphyseal union at the innominate and lower limb in a modern Portuguese skeletal sample, and age estimation in adolescent and young adult male and female skeletons. *American Journal of Physical Anthropology* 135: 161-170.

Cardoso HFV. 2008b. Age estimation of adolescent and young adult male and female skeletons II, epiphyseal union at the upper limb and scapular girdle in a modern Portuguese skeletal sample. *American Journal of Physical Anthropology* 137: 97-105.

Cardoso HFV. 2008c. Sample-specific (universal) metric approaches for determining the sex of immature human skeletal remains using permanent tooth dimensions. *Journal of Archaeological Science* 35: 158-168.

Cardoso HFV. 2009. Accuracy of developing tooth length as an estimate of age in human skeletal remains – the permanent dentition. *The American Journal of Forensic Medicine and Pathology* 30: 127-133.

Cardoso HFV, Garcia S. 2009. The Not-So-Dark Ages: ecology for human growth in medieval and early twentieth century Portugal as inferred from skeletal growth profiles. *American Journal of Physical Anthropology* 138: 136-147.

Cardoso HFV, Gomes JEA. 2008. Trends in adult stature of peoples who inhabited the modern Portuguese territory from the Mesolithic to the late 20th century. *International Journal of Osteoarchaeology* 19: 711-725.

Cardoso HFV, Saunders SR. 2008. Two arch criteria of the ilium for sex determination of immature skeletal remains: a test of their accuracy and an assessment of intra- and inter-observer error. *Forensic Science International* 178: 24-29.

Carlson KJ, Grine FE, Pearson OM. 2007. Robusticity and sexual dimorphism in the postcranium of modern hunter-gatherers from Australia. *American Journal of Physical Anthropology* 134: 9-23.

Caspari R. 2009. 1918: three perspectives on race and human variation. *American Journal of Physical Anthropology* 139: 5-15.

Castillo RF, Ubelaker DH, Djorojevic M. 2012. Age estimation through histological study of trabecular volume and cortical bone width of the iliac crest. *Science & Justice* 52: 177-180.

Chamberlain A. 2000. Problems and prospects in palaeodemography. In *Human Osteology in Archaeology and Forensic Science,* Cox M, Mays S (eds.). Greenwich Medical Media Ltd.: London; 101-115.

Chamberlain A. 2006. *Demography in Archaeology.* Cambridge University Press: Cambridge.

Charles DK, Condon K, Cheverud JM, Buikstra JE. 1986. Cementum annulations and age determination in *Homo sapiens* I. Tooth variability and observer error. *American Journal of Physical Anthropology* 71: 311-320.

Chikhi L, Destro-Bisol G, Pascall V, Baravelli V, Dobosz M, Barbujani G. 1998. Clinal variation in the nuclear DNA of Europeans. *Human Biology* 70: 643-657.

Chow W-H, Dong LM, Devesa SS. 2010. Epidemiology and risk factors for kidney cancer. *Nature Reviews Urology* 7: 245-257.

Coale AJ, Demeny P. 1983. *Regional Model Life Tables and Stable Populations*. Academic Press: Toronto.

Cobb WM. 1955. The age of incidence of suture closure. (Abstract) *American Journal of Physical Anthropology* 13: 394.

Cohen MM. 1993. Sutural biology and the correlates of craniosynostosis. *American Journal of Medical Genetics* 47: 581-616.

Coleman WH. 1969. Sex differences in the growth of the human bony pelvis. *American Journal of Physical Anthropology* 31: 125-152.

Conceição ELN, Cardoso HFV. 2011. Environmental effects on skeletal versus dental development II: further testing of a basic assumption in human osteological research. *American Journal of Physical Anthropology* 144: 463-470.

Coon CC. 1963. *The Origin of Races*. Knopf: New York.

Coqueugniot H, Weaver TD. 2007. Brief communication: infracranial maturation in the skeletal collection from Coimbra, Portugal: new aging standards for epiphyseal union. *American Journal of Physical Anthropology* 134: 424-437.

Correia H, Balseiro S, De Areia M. 2005. Sexual dimorphism in the human pelvis: testing a new hypothesis. *HOMO – Journal of Comparative Human Biology* 56: 153-160.

Cowal LS, Pastor RF. 2008. Dimensional variation in the proximal ulna: evaluation of a metric method for sex assessment. *American Journal of Physical Anthropology* 135: 469-478.

Cox M. 1996. *Life and Death in Spitalfields, 1700 to 1850*. Council for British Archaeology: York.

Cox M. 2000. Ageing adults from the skeleton. In *Human Osteology in Archaeology and Forensic Science,* Cox M, Mays S (eds.). Greenwich Medical Media Ltd.: London; 61-81.

Cox M, Scott A. 1992. Evaluation of the obstetric significance of some pelvic characters in an 18th century British sample of known parity status. *American Journal of Physical Anthropology* 89: 431-440.

Cunha E, van Vark GN. 1991. The construction of sex discriminant functions from a large collection of skulls of known sex. *International Journal of Anthropology* 6: 53-66.

Curate F, Albuquerque A, Cunha EM. In press. Age-at-death estimation using bone densitometry: testing the Fernández Castillo and López Ruiz method in two documented skeletal samples from Portugal. *Forensic Science International* , http://dx.doi.org/10.1016/jforsciint.2012.12.002 (accessed online 05/08/2013).

Crews DE. 2003. *Human Senescence – Evolutionary and Biocultural Perspectives*. Cambridge University Press: Cambridge.

Crosse DJM. 2011. Human lifespan: what determines the intrinsic length of human lives? *Reviews in Clinical Gerontology* 21: 193-200.

Danielsson L. 1964. Incidence and prognosis of coxarthrosis. *Acta Orthopaedica Scandinavica, Supplementum* 66: 13-18.

Danielsson L, Hernborg J. 1970. Clinical and roentgenologic study of knee joints with osteophytes. *Clinical Orthopaedics & Related Research* 69: 302-312.

Davies SC, Fowler T, McKee M. 2012. On the state of public health in England. *The Lancet* 380: 1793-1795.

Dayal MR, Kegley ADT, Strkalj G, Bidmos MA, Kuykendall KL. 2009. The history and composition of the Raymond A. Dart collection of human skeletons at the University of the Witwatersrand, Johannesburg, South Africa. *American Journal of Physical Anthropology* 140: 324-335.

Dayal MR, Spocter MA, Bidmos MA. 2008. An assessment of sex using the skull of black South Africans by discriminant function analysis. *HOMO – Journal of Comparative Human Biology* 59: 209-221.

Derry DE. 1909. Note on the innominate bone as a factor in the determination of sex: with special reference to the sulcus preauricularis. *Journal of Anatomy and Physiology* 43: 266-276.

de Magalhães JP, Wuttke D, Wood SH, Plank M, Vora C. 2012. Genome-enviroment interactions that modulate aging: powerful targets for drug discovery. *Pharmacological Reviews* 64: 88-101.

De Mendonça MC. 2000. Estimation of height from the length of long bones in a Portuguese adult population. *American Journal of Physical Anthropology* 112: 39-48.

De Villiers H. 1968. Sexual dimorphism of the South African Bantu-speaking negro. *South African Journal of Science* 64: 118-124.

Dias N. 1989. Séries de crânes et armée de squelettes: les collections anthropologiques en France dans la seconde moitié du XIXe siècle. *Bulletins et Mémoires de la Société d'anthropologie de Paris* 1: 203-230.

DiBennardo R, Taylor JV. 1982. Classification and misclassification in sexing the black femur by discriminant function analysis. *American Journal of Physical Anthropology* 58: 145-151.

DiBennardo R, Taylor JV. 1983. Multiple discriminant function analysis of sex and race in the postcranial skeleton. *American Journal of Physical Anthropology* 61: 305-314.

Djurić M, Djonić D, Nikolić S, Popović D, Marinković J. 2007. Evaluation of the Suchey-Brooks method for aging skeletons in the Balkans. *Journal of Forensic Sciences* 52: 21-23.

Dreier FG. 1994. Age-at-death estimates for the protohistoric Arikara using molar attrition rates: a new quantification method. *International Journal of Osteoarchaeology* 4: 137-147.

Dudar JC. 1993. Identification of rib number and assessment of intercostals variation at the sternal rib end. *Journal of Forensic Sciences* 38: 788-797.

Đurić M, Rakočevič Z, Đonić D. 2005. The reliability of sex determination of skeletons from forensic context in the Balkans. *Forensic Science International* 147: 159-164.

Dwight T. 1894. The range and significance of variations in the human skeleton. *The Boston Medical and Surgical Journal* 131: 97-101.

Easton J. 1799. *Human Longevity: Recording the Name, Age, Place of Residence, and Year of the Decease of 1712 Persons Who Attained a Century and Upwards from AD 66 to 1799, Comprising A Period of 1733 Years.* Printed by James Easton: Salisbury.

Edwards BK, Ward E, Kohler BA, Eheman C, Zauber AG, Anderson RN, Jemal A, Schymura MJ, Lansdorp-Vogelaar I, Seeff LC, van Ballegooijen M, Goede L, Ries LAG. 2010. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* 116: 544-573.

Eisdorfer C, Mullner R, Cohen D. 1989. The role of the hospital in the care of older persons. In Ory MG and Bond K (eds.), *Aging and Health Care – Social Science and Policy Perspectives*. Routledge: London; 99-115.

Eliopoulos C, Lagia A, Manolis S. 2007. A modern, documented human skeletal collection from Greece. *HOMO – Journal of Comparative Human Biology* 58: 221-228.

Ericksen MF. 1982. How "representative" is the Terry Collection? Evidence from the proximal femur. *American Journal of Physical Anthropology* 59: 345-350.

Eveleth PB, Bowers EJ, Schall JI. 1979. Secular change in growth of Philadelphia black adolescents. *Human Biology* 51: 213-228.

Falys CG, Lewis ME. 2011. Proposing a way forward: a review of standardisation in the use of age categories and ageing techniques in osteological analysis (2004-2009). *International Journal of Osteoarchaeology* 21: 704-716.

Falys CG, Schutkowski H, Weston DA. 2006. Auricular surface aging: worse than expected? A test of the revised method on a documented historic skeletal assemblage. *American Journal of Physical Anthropology* 130: 508-513.

Ferré C. 1876. Sur l'atrophie sénile symétrique de pariétaux. *Bulletins de la Société d'Anthropologie de Paris* 11: 423-426.

Finch CE, Kirkwood TBL. 2000. *Chance, Development, and Aging*. Oxford University Press: Oxford.

Finkel T, Holbrook NJ. 2000. Oxidants, oxidative stress and the biology of ageing. *Nature* 408: 239-247.

FitzPatrick DR. 2013. Filling in the gaps in cranial suture biology. *Nature Genetics* 45: 231-232.

Franklin D, Cardini A. 2007. Mandibular morphology as an indicator of human subadult age: interlandmark approaches. *Journal of Forensic Sciences* 52: 1015-1019.

Franklin D, Freedman L, Milne N. 2005. Sexual dimorphism and discriminant function sexing in indigenous South African crania. *HOMO – Journal of Comparative Human Biology* 55: 213-228.

Franklin D, Oxnard CE, O'Higgins P, Dadour I. 2007. Sexual dimorphism in the subadult mandible: quantification using geometric morphometrics. *Journal of Forensic Sciences* 52: 6-10.

Frayer DW, Wolpoff MH. 1985. Sexual dimorphism. *Annual Review of Anthropology* 14: 429-473.

Galera V, Ubelaker DH, Hayek LC. 1995. Interobserver error in macroscopic methods of estimating age-at-death from the human skeleton. *International Journal of Anthropology* 10: 229-239.

Garn SM. 1987. The secular trend in size and maturational timing and its implications for nutritional assessment. *The Journal of Nutrition* 117: 817-823.

Garvin HM, Passalacqua NV. 2012. Current practices by forensic anthropologists in adult skeletal age estimation. *Journal of Forensic Sciences* 57: 427-433.

Gauthier J, Schutkowski H. 2013. Assessing the application of tooth cementum annulations relative to macroscopic aging techniques in an archeological sample. *HOMO – Journal of Comparative Human Biology* 64: 42-57.

Genoves, ST. 1960. Revaluation of age, stature and sex of the Tepexpan remains, Mexico. *American Journal of Physical Anthropology* 18: 205-217.

Gilbert BM. 1973. Misapplication to females of the standard for aging the male *os pubis*. *American Journal of Physical Anthropology* 38: 39-40.

Gilbert BM, McKern TW. 1973. A method for aging the female *Os pubis*. *American Journal of Physical Anthropology* 38: 31-38.

Giles E, Elliot O. 1963. Sex determination by discriminant function analysis of crania. *American Journal of Physical Anthropology* 21: 53-68.

Gill GW, Rhine S. (eds.) 1990. *Skeletal Attribution of Race: Methods for Forensic Anthropology*. Anthropological Papers, No. 4. Maxwell Museum of Anthropology: Albuquerque.

Gillett RM. 1991. Determination of age-at-death in human skeletal remains: a comparison of two techniques. *International Journal of Anthropology* 6: 179-189.

Gilmore CC, Grote MN. 2012. Estimating age from adult occlusal wear: a modification of the Miles method. *American Journal of Physical Anthropology* 149: 181-192.

Giraudi R, Fissore F, Giacobini G. 1984. The collection of human skulls and postcranial skeletons at the Department of Human Anatomy of the University of Torino (Italy). *American Journal of Physical Anthropology* 65: 105-107.

Glowacki J, Zhou S. 2007. Extrinsic and intrinsic mechanisms of human skeletal aging. *Orthopaedic Journal at Harvard Medical School* 9: 80-82.

Godde K, Hens SM. 2012. Age-at-death estimation in an Italian historical sample: a test of the Suchey-Brooks and transition analysis methods. *American Journal of Physical Anthropology* 149: 259-265.

Gonzalez PN, Perez SI, Bernal V. 2010. Ontogeny of robusticity of craniofacial traits in modern humans: a study of South American populations. *American Journal of Physical Anthropology* 142: 367-379.

Gonzalez PN, Bernal V, Perez SI. 2011. Analysis of sexual dimorphism of craniofacial traits using geometric morphometric techniques. *International Journal of Osteoarchaeology* 21: 87-91.

Gooney D. 2012. Treatment of the dead in Iron Age Orkney. An examination of the burials from the cemetery at Knowe of Skea, Westray. Presentation at the 14th Annual Conference of the British Association for Biological Anthropology and Osteoarchaeology, September 2012, Bournemouth.

Gowland R. 2007. Age, ageism and osteological bias: the evidence from late Roman Britain. *Journal of Roman Archaeology* (supplementary series) 65: 153-169.

Gowland R, Chamberlain AT. 2005. Detecting plague: palaeodemographic characterisation of a catastrophic death assemblage. *Antiquity* 79: 146-157.

Gravlee CC. 2009. How race becomes biology: embodiment of social inequality. *American Journal of Physical Anthropology* 139: 47-57.

Gray JP, Wolfe LD. 1980. Height and sexual dimorphism of stature among human societies. *American Journal of Physical Anthropology* 53: 441-456.

Griffin RC, Chamberlain AT, Hotz G, Penkman KEH, Collins MJ. 2009. Age estimation of archaeological remains using amino acid racemisation in dental enamel: a comparison of morphological, biochemical, and known ages-at-death. *American Journal of Physical Anthropology* 140: 244-252.

Grundy E, Glaser K. 2000. Socio-demographic differences in the onset and progression of disability in early old age: a longitudinal study. *Age and Ageing* 29: 149-157.

Gu D, Kelly TN, Hixson JE, Chen J, Liu D, Chen J-C, Rao DC, Mu J, Ma J, Jaquish CE, Rice TK, Gu C, Hamm LL, Whelton PK, He J. 2010. Genetic variants in the renin-angiotensin-aldosterone system and salt-sensitivity of blood pressure. *Journal of Hypertension* 28: 1210-1220.

Gustafson G. 1950. Age determination on teeth. *Journal of the American Dental Association*. 41: 45-54.

Handley LJL, Manica A, Goudet J, Balloux F. 2007. Going the distance: human population genetics in a clinal world. *Trends in Genetics* 23: 432-439.

Hanna RE, Washburn SL. 1953. The determination of the sex of skeletons as illustrated by a study of the Eskimo pelvis. *Human Biology* 25: 21-27.

Hannan MT, Felson DT, Dawson-Hughes B, Tucker KL, Cupples LA, Wilson PWF, Kiel DP. 2000. Risk factors for longitudinal bone loss in elderly men and women: the Framingham Osteoporosis Study. *Journal of Bone and Mineral Research* 15: 710-720.

Hartnett KM. 2010a. Analysis of age-at-death estimation using data from a new, modern autopsy sample – part I: pubic bone. *Journal of Forensic Sciences* 55: 1145-1151.

Hartnett KM. 2010b. Analysis of age-at-death using data from a new, modern autopsy sample – part II: sternal end of the fourth rib. *Journal of Forensic Sciences* 55: 1152-1156.

Hens SM, Belcastro MG. 2012. Auricular surface aging: a blind test of the revised method on historic Italians from Sardinia. *Forensic Science International* 214: 209.e1-209.e5.

Hens SM, Rastelli E, Belcastro G. 2008. Age estimation from the human os coxa: a test on a documented Italian collection. *Journal of Forensic Sciences* 53: 1040-1043.

Hernborg J, Nilsson BE. 1973. The relationship between osteophytes in the knee joint, osteoarthritis and aging. *Acta Orthopaedica* 44: 69-74.

Hershkovitz I, Latimer B, Dutour O, Jellema LM, Wish-Baratz S, Rothschild C, Rothschild BM. 1997. Why do we fail in aging the skull from the sagittal suture? *American Journal of Physical Anthropology* 103: 393-399.

Heuzé Y, Braga J. 2008. Application of non-adult Bayesian dental age assessment methods to skeletal remains: the Spitalfields collection. *Journal of Archaeological Science* 35: 368-375.

Heuzé Y, Cardoso HFV. 2008. Testing the quality of nonadult Bayesian dental age assessment methods to juvenile skeletal remains: the Lisbon Collection children and secular trend effects. *American Journal of Physical Anthropology* 135: 275-283.

Hippocrates. 400 BCE. *On Injuries of the Head.*

HM Government. 2008. Preparing for our Ageing Society: A Discussion Paper.

Hooton EA. 1918. On certain Eskimoid characters in Icelandic skulls. *American Journal of Physical Anthropology* 1: 53-76.

Hopkins K. 1966. On the probable age structure of the Roman population. *Population Studies* 20: 245-264.

Hoppa RD. 2000. Population variation in osteological aging criteria: an example from the pubic symphysis. *American Journal of Physical Anthropology* 111: 185-191.

Hoppa RD, Vaupel JW. 2002. The Rostock Manifesto for paleodemography: the way from stage to age. In *Paleodemography – Age Distributions from Skeletal Samples*, Hoppa RD, Vaupel JW (eds.). Cambridge University Press: Cambridge; 1-8.

Hosegood V, Vanneste A-M, Timaeus IM. 2004. Levels and causes of adult mortality in rural South Africa: the impact of AIDS. *AIDS* 18: 663-671.

Howell N. 1976. Toward a uniformitarian theory of human paleodemography. In *The Demographic Evolution of Human Populations*, Ward RH, Weiss KM (eds.). Academic Press: New York; 25-40.

Hrdlička A. 1918. Physical anthropology: its scope and aims; its history and present status in America. *American Journal of Physical Anthropology* 1: 3-23.

Hughes KA. 2010. Mutation and the evolution of ageing: from biometrics to system genetics. *Philosophical Transactions of the Royal Society* 365: 1273-1279.

Humphrey LT, Scheuer L. 2006. Age of closure of the foramen of Huschke: an osteological study. *International Journal of Osteoarchaeology* 16: 47-60.

Humphry GM. 1890. Senile hypertrophy and senile atrophy of the skull. *Journal of Anatomy* 24: 598.

Hunt DR, Albanese J. 2005. History and demographic composition of the Robert J. Terry Anatomical Collection. *American Journal of Physical Anthropology* 127: 406-417.

Hyde Z, Flicker L, McCaul K, Almeida OP, Hankey GJ, Chubb SAP, Yeap BB. 2012. Associations between testosterone levels and incident prostate, lung, and colorectal cancer. A population-based study. *Cancer Epidemiology, Biomarkers & Prevention* 21: 1319-1329.

Hynes J. 1995. The oldest old in pre-industrial Britain: centenarians before 1800 – fact or fiction? In *Exceptional Longevity: From Prehistory to the Present*, Odense Monographs on Population Aging 2, Jeune B, Vaupel JW (eds.). Odense University Press: Odense; 75-92.

İşcan MY, Loth SR. 1989. Osteological manifestations of age in the adult. In *Reconstruction of Life from the Skeleton*, İşcan MY, Kennedy KAR (eds.). Alan R. Liss, Inc.: New York; 23-40.

İşcan MY, Loth SR. 1993. Casts of age phases from the sternal end of the rib for white males and females. France Casts, Bellvue Co.

İşcan MY, Loth SR, Wright RK. 1984. Age estimation from the rib by phase analysis: white males. *Journal of Forensic Sciences* 29: 1094-1104.

İşcan MY, Loth SR, Wright RK. 1984b. Metamorphosis at the sternal rib end: a new method to estimate age-at-death in white males. *American Journal of Physical Anthropology* 65: 147-156.

İşcan MY, Loth SR, Wright RK. 1985. Age estimation from the rib by phase analysis: white females. *Journal of Forensic Sciences* 30: 853-863.

İşcan MY, Loth SR, Wright RK. 1987. Racial variation in the sternal extremity of the rib and its effect on age distribution. *Journal of Forensic Sciences* 32: 452-466.

İşcan MY, Shihai D. 1995. Sexual dimorphism in the Chinese femur. *Forensic Science International* 74: 79-87.

İşcan MY, Yoshino M, Kato S. 1995. Sexual dimorphism in modern Japanese crania. *American Journal of Human Biology* 7: 459-464.

Jackes MK. 1985. Pubic symphysis age distributions. *American Journal of Physical Anthropology* 68: 281-299.

Jackes MK. 2000. Building the bases for paleodemographic analysis: adult age determination. In *Biological Anthropology of the Human Skeleton*, Katzenberg MA, Saunders SR (eds.). John Wiley and Sons: New York; 407-456.

Jantz RL, Meadows Jantz L. 2000. Secular change in craniofacial morphology. *American Journal of Human Biology* 12: 327-338.

Jordana X, Isidro A, Malgosa A. 2010. Interpreting diachronic osteological variation at the medieval necropolis of the Sant Pere churches (Terrassa, Spain). *International Journal of Osteoarchaeology* 20: 670-692.

Joshi PS, Chougule MS, Agrawal GP. 2010. Comparison of polarizing & phase contrast microscopy for estimation of age based on cemental annulations. *Indian Journal of Forensic Odontology* 3: 87-95.

Kahn K, Tollman SM, Garenne M, Gear JSS. 1999. Who dies from what? Determining cause of death in South Africa’s rural north-east. *Tropical Medicine and International Health* 4: 433-441.

Kajanoja P. 1966. Sex determination of Finnish crania by discriminant function analysis. *American Journal of Physical Anthropology* 24: 29-34.

Kalichman L, Livshits G, Kobyliansky E. 2006. Association between morbidity and skeletal biomarkers of biological aging. *Human Biology* 78: 77-88.

Kaplan GA, Seeman TE, Cohen RD, Knudsen LP, Guralnik J. 1987. Mortality among the elderly in the Alameda County Study: behavioral and demographic risk factors. *American Journal of Public Health* 77: 307-312.

Karasik D, Demissie S, Cupples LA, Kiel DP. 2005. Disentangling the genetic determinants of human aging: biological age as an alternative to the use of survival measures. *Journal of Gerontology* 60A: 574-587.

Katz D, Suchey JM. 1986. Age determination of the male os pubis. *American Journal of Physical Anthropology* 69: 427-435.

Katz D, Suchey JM. 1989. Race differences in pubic symphyseal aging patterns in the male. *American Journal of Physical Anthropology* 80: 167-172.

Keen JA. 1950. A study of the differences between male and female skulls. *American Journal of Physical Anthropology* 8: 65-79.

Keita SOY, Kittles RA. 1997. The persistence of racial thinking and the myth of racial divergence. *American Anthropologist* 99: 534-544.

Kelley MA. 1978. Phenice’s visual sexing technique for the os pubis: a critique. *American Journal of Physical Anthropology* 48: 121-122.

Kemkes-Grottenthaler A. 2002. Aging through the ages: historical perspectives on age indicator methods. In *Paleodemography – Age Distributions from Skeletal Samples*, Hoppa RD, Vaupel JW (eds.). Cambridge University Press: Cambridge; 48-72.

Kemkes-Grottenthaler A. 2005. Sex determination by discriminant analysis: an evaluation of the reliability of patella measurements. *Forensic Science International* 147: 129-133.

Keough N, L'Abbé EN, Steyn M. 2009. The evaluation of age-related histomorphometric variables in a cadaver sample of lower socioeconomic status: implications for estimating age-at-death. *Forensic Science International* 191: 114.e1-114.e6.

Kerley ER. 1965. The microscopic determination of age in human bone. *American Journal of Physical Anthropology* 23: 149-163.

Key CA, Aiello LC, Molleson T. 1994. Cranial suture closure and its implications for age estimation. *International Journal of Osteoarchaeology* 4: 193-207.

Kimmerle EH, Jantz RL, Konigsberg LW, Baraybar JP. 2008. Skeletal estimation and identification in American and East European populations. *Journal of Forensic Sciences* 53: 524-532.

Kirkwood TBL. 1997. What is the relationship between osteoarthritis and ageing? *Baillière's Clinical Rheumatology* 11: 683-694.

Kirkwood TBL, Holliday R. 1986. Ageing as a consequence of natural selection. In *The Biology of Human Ageing*, Society for the Study of Human Biology Symposium 25, Bittles AH, Collins KJ (eds.). Cambridge University Press: Cambridge; 1-16.

Kjæraard T. 1995. Alleged Danish centenarians before 1800. In *Exceptional Longevity: From Prehistory to the Present*, Odense Monographs on Population Aging 2, Jeune B, Vaupel JW (eds.). Odense University Press: Odense; 47-54.

Klepinger LL, Katz D, Micozzi MS, Carroll L. 1992.Evaluation of case methods for estimating age from the *os pubis*. *Journal of Forensic Sciences* 37: 763-770.

Kobyliansky E, Livshits G, Pavlovsky O. 1995. Population biology of human aging: methods of assessment and sex variation. *Human Biology* 67: 87-109.

Komar DA, Grivas C. 2008. Manufactured populations: what do contemporary reference skeletal collections represent? A comparative study using the Maxwell Museum Documented Collection. *American Journal of Physical Anthropology* 137: 224-233.

Konigsberg LW, Frankenberg SR. 1992. Estimation of age structure in anthropological demography. *American Journal of Physical Anthropology* 89: 235-256.

Konigsberg LW, Frankenberg SR. 1994. Paleodemography: "not quite dead". *Evolutionary Anthropology* 3: 92-105.

Konigsberg LW, Hens SM. 1998. Use of ordinal categorical variables in skeletal assessment of sex from the cranium. *American Journal of Physical Anthropology* 107: 97-112.

Konigsberg LW, Herrmann NP. 2002. Markov chain Monte Carlo estimation of hazard model parameters in paleodemography. In *Paleodemography – Age Distributions from Skeletal Samples*, Hoppa RD, Vaupel JW (eds.). Cambridge University Press: Cambridge; 222-242.

Konigsberg LW, Herrmann NP, Wescott DJ. 2002. Commentary on McBride et al. "Bootstrap methods for sex determination from the os coxae using the ID3 algorithm". *Journal of Forensic Science* 47: 424-426.

Krall EA, Dawson-Hughes B. 1993. Heritable and life-style determinants of bone mineral density. *Journal of Bone and Mineral Research* 8: 1-9.

Kunos CA, Simpson SW, Russell KF, Hershkovitz I. 1999. First rib metamorphosis: its possible utility for human age-at-death estimation. *American Journal of Physical Anthropology* 110: 303-323.

Kvaal SI, Sellevold BJ, Solheim T. 1994. A comparison of different non-destructive methods of age estimation in skeletal material. *International Journal of Osteoarchaeology* 4: 363-370.

L'Abbé EN, Loots M, Meiring JH. 2005. The Pretoria bone collection: a modern South African skeletal sample. *HOMO – Journal of Comparative Human Biology* 56: 197-205.

Landi F, Russo A, Cesari M, Pahor M, Bernabei R, Onder G. 2007. HDL-cholesterol and physical performance: results from the ageing and longevity study in the sirente geographic area (*ilSIRENTE Study*). *Age and Ageing* 36: 514-520.

Lasker GW. 1969. Human biological adaptability: the ecological approach in physical anthropology. *Science* 166: 1480-1486.

LeBoff MS, Glowacki J. 1999. Sex steroids, bone, and aging. In *The Aging Skeleton*, Rosen CJ, Glowacki J, Bilezikian JP (eds.). Academic Press: London; 159-174.

Letterman GS. 1941. The greater sciatic notch in American whites and negroes. *American Journal of Physical Anthropology* 28: 99-116.

Lewis ME. 2002. Impact of industrialization: comparative study of child health in four sites from medieval and postmedieval England (A.D. 850-1859). *American Journal of Physical Anthropology* 119: 211-223.

Lewis ME. 2007. *The Bioarchaeology of Children*. Cambridge University Press: Cambridge.

Lewis ME, Gowland R. 2007. Brief and precarious lives: infant mortality in contrasting sites from medieval and post-medieval England (AD 850-1859). *American Journal of Physical Anthropology* 134: 117-129.

Lewontin RC. 1972. The apportionment of human diversity. *Evolutionary Biology* 6: 381-398.

Listi GA, Manhein MH. 2012. The use of vertebral osteoarthritis and osteophytosis in age estimation. *Journal of Forensic Sciences* 57: 1537-1540.

Litten J. 2002. *The English Way of Death: The Common Funeral Since 1450.* Robert Hale: London.

Litwin H, Shiovitz-Ezra S. 2006. The association between activity and wellbeing in later life: what really matters? *Ageing & Society* 26: 225-242.

Liversidge HM. 1994. Accuracy of age estimation from developing teeth of a population of known age (0-5.4 years). *International Journal of Osteoarchaeology* 4: 37-45.

Liversidge HM, Molleson TI. 1999. Deciduous tooth size and morphogenetic fields in children from Christ Church, Spitalfields. *Annals of Oral Biology* 44: 7-13.

Livshits G, Pavlovsky O, Kobyliansky E. 1996. Population biology of human aging: segregation analysis of bone age characteristics. *Human Biology* 68: 539-554.

Lorenzini A, Stamato T, Sell C. 2011. The disposable soma theory revisited. Time as a resource in the theories of aging. *Cell Cycle* 10: 3853-3856.

Loth SR. 1995. Age assessment of the Spitalfields cemetery population by rib phase analysis. *American Journal of Human Biology* 7: 465-471.

Loth SR, Henneberg M. 1996. Mandibular ramus flexure: a new morphologic indicator of sexual dimorphism in the human skeleton. *American Journal of Physical Anthropology* 99: 473-485.

Loth SR, Henneberg M. 2001. Sexually dimorphic mandibular morphology in the first few years of life. *American Journal of Physical Anthropology* 115: 179-186.

Loth SR, İşcan MY. 1989. Morphological assessment of age in the adult: the thoracic region. In *Age Markers in the Human Skeleton*, İşcan MY (ed.). Springfield: Illinois; 105-135.

Lovejoy CO, Meindl RS, Mensforth RP, Barton TJ. 1985. Multifactorial determination of skeletal age-at-death: a method of blind tests of its accuracy. *American Journal of Physical Anthropology*. 68: 1-14.

Lovejoy CO, Meindl RS, Pryzbeck TR, Barton TS, Heiple KG, Kotting D. 1977. Paleodemography of the Libben site, Ottawa County, Ohio. *Science* 198: 291-293.

Lovejoy CO, Meindl RS, Pryzbeck TR, Mensforth RP. 1985b. Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age-at-death. *American Journal of Physical Anthropology* 68: 15-28.

Lovell NC. 1989. Test of Phenice's technique for determining sex from the os pubis. *American Journal of Physical Anthropology* 79: 117-120.

Luboga S 2000. Supernumerary lumbar verebrae in human skeletons at the Galloway Osteological Collection of Makerere University, Kampala. East African Medical Journal 77: 16-9.

Lungmus EK. 2009. An examination of error in the application of pubic aging techniques. *Unpublished MA thesis, University of Montana, USA.*

MacLaughlin SM, Bruce MF. 1986. The sciatic notch/acetabular index as a discriminator of sex in European skeletal remains. *Journal of Forensic Sciences* 31: 1380-1390.

MacLaughlin SM, Bruce MF. 1986b. Population variation in sexual dimorphism in the human innominate. *Human Evolution* 1: 221-231.

MacLaughlin SM, Bruce MF. 1990. The accuracy of sex identification in European skeletal remains using the Phenice characters. *Journal of Forensic Sciences* 35: 1384-1392.

Martrille L, Ubelaker DH, Cattaneao C, Seguret F, Tremblay M, Baccino E. 2007. Comparison of four skeletal methods for the estimation of age-at-death on white and black adults. *Journal of Forensic Sciences* 52: 302-307.

Masset C. 1989. Age estimation on the basis of cranial sutures. In *Age Markers in the Human Skeleton*, İşcan MY (ed.). Springfield: Illinois; 71-103.

Masters PM, Zimmerman MR. 1978. Age determination of an Alaskan mummy: morphological and biochemical correlation. *Science* 201: 811-812.

Mastrangelo P, De Luca S, Alemán I, Botella MC. 2011. Sex assessment from the carpals bones: discriminant function analysis in a 20th century Spanish sample. *Forensic Science International* 206: 216.e1-216.e10.

Mathers CD, Boerma T, Ma Fat D. 2009. Global and regional causes of death. *British Medical Bulletin* 92: 7-32.

Matos V. 2009. Broken ribs: paleopathological analysis of costal fractures in the human identified skeletal collection from the Museu Bocage, Lisbon, Portugal (late 19th to middle 20th centuries). *American Journal of Physical Anthropology* 140: 25-38.

Matos V, Santos AL. 2006. On the trail of pulmonary tuberculosis based on rib lesions: results from the human identified skeletal collection from the Museu Bocage (Lisbon, Portugal). *American Journal of Physical Anthropology* 130: 190-200.

Mays S. 2000. Age-dependent cortical bone loss in women from 18th and early 19th century London. *American Journal of Physical Anthropology* 112: 349-361.

Mays S. 2001. Effects of age and occupation on cortical bone in a group of 18th-19th century British men. *American Journal of Physical Anthropology* 116: 34-44.

Mays S. 2002. The relationship between molar wear and age in an early 19th century AD archaeological human skeletal series of documented age-at-death. *Journal of Archaeological Science* 29: 861-871.

Mays S, Cox M. 2000. Sex determination in skeletal remains. In *Human Osteology: In Archaeology and Forensic Science*, Cox M, Mays S (eds.). Cambridge University Press: Cambridge; 117-130.

Mazess RB, Mathisen RW. 1982. Lack of unusual longevity in Vilcabamba, Ecuador. *Human Biology* 54: 517-524.

McCracken M, Olsen M, Chen, Jr. MS, Jemal A, Thun M, Cokkinides V. 2009. Cancer incidence, mortality, and associated risk factors among Asian Americans of Chinese, Filipino, Vietnamese, Korean and Japanese Ethnicities. *CA: A Cancer Journal of Clinicians* 57: 190-205.

McGee DL, Liao Y, Cao G, Cooper RS. 1999. Self-reported health status and mortality in a multiethnic US cohort. *American Journal of Epidemiology* 149: 41-46.

McKern TW, Stewart SD. 1957. *Skeletal Age Changes in Young American Males. Analyzed from the Skeletal Standpoint of Age Identification.* US Army Quartermaster Research and Development Center: Natick, MA.

Megyesi MS, Ubelaker DH, Sauer NJ. 2006. Test of the Lamendin aging method on two historic skeletal samples. *American Journal of Physical Anthropology* 131: 363-367.

Meindl RS, Lovejoy CO. 1985. Ectocranial suture closure: a revised method for the determination of skeletal age-at-death based on the lateral-anterior sutures. *American Journal of Physical Anthropology* 68: 57-66.

Meindl RS, Lovejoy CO, Mensforth RP. 1983. Skeletal age-at-death: accuracy of determination and implications for human demography. *Human Biology* 55: 73-87.

Meindl RS, Lovejoy CO, Mensforth RP, Walker RA. 1985. A revised method of age determination using the os pubis, with a review and tests of accuracy of other current methods of pubic symphyseal aging. *American Journal of Physical Anthropology* 68: 29-45.

Meindl RS, Lovejoy CO, Mensforth RP, Don Carlos L. 1985b. Accuracy and direction of error in the sexing of the skeleton: implications for paleodemography. *American Journal of Physical Anthropology* 68: 79-85.

Meindl RS, Russell KF, Lovejoy CO. 1990. Reliability of age-at-death in the Hamann-Todd Collection: validity of subselection procedures used in blind tests of the summary age technique. *American Journal of Physical Anthropology* 83: 349-357.

Meindl RS, Russell KF. 1998. Recent advances in method and theory in paleodemography. *Annual Review of Anthropology* 27: 375-399.

Mensforth RP. 1990. Paleodemography of the Carlston Annis (Bt-5) Late Archaic skeletal population. *American Journal of Physical Anthropology* 82: 81-99.

Mensforth RP, Lovejoy CO. 1985. Anatomical, physiological and epidemiological correlates of the aging process: a confirmation of multifactorial age determination in the Libben skeletal population. *American Journal of Physical Anthropology* 68: 87-106.

Milne N. 1990. Sexing of human hip bones. *Journal of Anatomy* 172: 221-226.

Milner G. 2007. Paleodemography and Osteological Age Estimation Summer School, University of Southern Denmark and Max Planck Institute for Demographic Research, August 2007. (*Personal communication.*)

Molleson TI. 1986. Skeletal age and palaeodemography. In *The Biology of Human Ageing*, Society for the Study of Human Biology Symposium 25, Bittles AH, Collins KJ (eds.). Cambridge University Press: Cambridge; 95-118.

Molleson T. 1995. Rates of ageing in the eighteenth century. In *Grave Reflections: Portraying the Past Through Cemetery Studies*, Saunders SR, Herring A (eds.). Canadian Scholars Press: Toronto; 199-222.

Molleson T, Cox M. 1993. *The Spitalfields project. Volume 2 – the anthropology. The middling sort.* CBA research report 86. York: Council for British Archaeology.

Molnar S. 1971. Human tooth wear, tooth function and cultural variability. *American Journal of Physical Anthropology* 34: 175-190.

Montagu JD. 1994. Length of life in the ancient world: a controlled study. *Journal of the Royal Society of Medicine* 87: 25-26.

Moore LG, Niermeyer S, Zamudio S. 1998. Human adaptation to high altitude: regional and life-cycle perspectives. *Yearbook of Physical Anthropology* 41: 25-64.

Mozaffarian D, Kamineni A, Carnethon M, Djoussé L, Mukamal KJ, Siscovick D. 2009. Lifestyle risk factors and new-onset diabetes mellitus in older adults. *Archives of Internal Medicine* 169: 798-807.

Mulhern DM, Jones EB. 2005. Test of revised method of age estimation from the auricular surface of the ilium. *American Journal of Physical Anthropology*. 126: 61-65.

Munch TN, Rostgaard K, Rasmussen M-LH, Wohlfahrt J, Juhler M, Melbye M. 2012. Familial aggregation of congenital hydrocephalus in a nationwide cohort. *Brain* 135: 2409-2415.

Murphy TR. 1959. The changing pattern of dentine exposure in human tooth attrition. *American Journal of Physical Anthropology* 17: 167-176.

Murray KA, Murray T. 1991. A test of the auricular surface aging technique. *Journal of Forensic Sciences* 36: 1162-1169.

Myerowitz R, Hogikyan ND. 1986. Different mutations in Ashkenazi Jewish and non-Jewish French Canadians with Tay-Sachs disease. *Science* 232: 1646-1648.

Myles S, Lea RA, Ohashi J, Chambers GK, Weiss JG, Hardouin E, Engelken J, Macartney-Coxson DP, Eccles DA, Naka I, Kimura R, Inaoka T, Matsumara Y, Stoneking M. 2011. Testing the thrifty gene hypothesis: the Gly482Ser variant in *PPARGC1A* is associated with BMI in Tongans. *BMC Medical Genetics* 12: 1-10.

Nagaoka T, Hirata K. 2008. Demographic structure of skeletal populations in historic Japan: a new estimation of adult age-at-death distributions based on the auricular surface of the ilium. *Journal of Archaeological Science* 35: 1370-1377.

Nagaoka T, Hirata K, Yokota E, Matsu'ura S. 2006. Paleodemography of a medieval population in Japan: analysis of human skeletal remains from the Yuigahama-minami site. *American Journal of Physical Anthropology* 131: 1-14.

Nagaoka T, Sawada J, Hirata K. 2008. Did the Jomon people have a short lifespan? Evidence from the adult age-at-death estimation based on the auricular surface of the ilium. *Anthropological Science* 116: 161-169.

Nagar Y, Hershkovitz I. 2004. Interrelationship between various aging methods, and their relevance to palaeodemography. *Human Evolution* 19: 145-156.

Nathan H. 1962. Osteophytes of the vertebral column: an anatomical study of their development according to age, race, and sex with considerations as to their etiology and significance. *The Journal of Bone and Joint Surgery* 44: 243-268.

Nicholson C. 1945. The two main diameters at the brim of the female pelvis. *The Journal of Anatomy* 79: 131-135.

Nieminen P, Morgan NV, Fenwick AL, Parmanen S, Veistinen L, Mikkola ML, van der Spek PJ, Giraud A, Judd L, Arte S, Brueton LA, Wall SA, Mathijssen IMJ, Maher ER, Wilkie AOM, Kreiborg S, Thesleff I. 2011. Inactivation of IL11 signaling causes craniosynostosis, delayed tooth eruption, and supernumerary teeth. *The American Journal of Human Genetics* 89: 67-81.

Oettlé AC, Becker PJ, de Villiers E, Steyn M. 2009. The influence of age, sex, population group and dentition on the mandibular angle as measured on a South African sample. *American Journal of Physical Anthropology* 139: 505-511.

Oettlé AC, Pretorius E, Steyn M. 2005. Geometric morphometric analysis of mandibular ramus flexure. *American Journal of Physical Anthropology* 128: 623-629.

Oettlé AC, Steyn M. 2000. Age estimation from sternal ends of ribs by phase analysis in South African blacks. *Journal of Forensic Sciences* 45: 1071-1079.

Oliveira RN, Silva SFSM, Kawano A, Antunes JLF. 2006. Estimating age by tooth wear of prehistoric human remains in Brazilian archaeological sites. *International Journal of Osteoarchaeology* 16: 407-414.

ONS. 2011. *National Population Projections, 2010-Based Statistical Bulletin*. Office for National Statistics: London.

ONS, 2012. *Population Ageing in the UK and EU*. Office for National Statistics: London. Online video: http://www.youtube.com/watch?v=50tdLusm7-Y (accessed 23/03/2013).

Opperman LA. 2000. Cranial sutures as intramembranous bone growth sites. *Developmental Dynamics* 219: 472-485.

Ortner DJ. 2003. *Identification of Pathological Conditions in Human Skeletal Remains.* Academic Press: London.

Osborne DL, Simmons TL, Nawrocki SP. 2004. Reconsidering the auricular surface as an indicator of age-at-death. *Journal of Forensic Sciences* 49: 1-7.

Ostojić S, Pereza N, Kapović M. 2009. A current genetic and epigenetic view on human aging mechanisms. *Collegium Antropologicum* 33: 687-699.

Owers SK, Pastor RF. 2005. Analysis of quantitative methods for rib seriation using the Spitalfields documented skeletal collection. *American Journal of Physical Anthropology* 127: 210-218.

Paine RR. 1997. Uniformitarian models in osteological paleodemography. In *Integrating Archaeological Demography – Multidisciplinary Approaches to Prehistoric Population*, Paine RR (ed.). Occasional Paper No. 24, Center for Archaeological Investigations, Southern Illinois University: Carbondale; 191-204.

Parson FG. 1914. The characters of the English thigh-bone. *Journal of Anatomy and Physiology* 48: 238-267.

Parsons FG. 1915. The characters of the English thigh-bone: Part II: the difficulty of sexing. *Journal of Anatomy and Physiology* 49: 345-361.

Parsons FG, Box CR. 1905. The relation of the cranial sutures to age. *The Journal of the Anthropological Institute of Great Britain and Ireland* 35: 30-38.

Parsons FG, Keene L. 1919. Sexual differences in the skull. *Journal of Anatomy* 54: 58-65.

Passalacqua NV. 2010. The utility of the Samworth and Gowland age-at-death "look-up" tables in forensic anthropology. *Journal of Forensic Sciences* 55: 482-487.

Patriquin ML, Loth SR, Steyn M. 2003. Sexually dimorphic pelvic morphology in South African whites and blacks. *HOMO – Journal of Comparative Human Biology* 53: 255-262.

Patriquin ML, Steyn M, Loth SR. 2002. Metric assessment of race from the pelvis in South Africans. *Forensic Science International* 127: 104-113.

Patriquin ML, Steyn M, Loth SR. 2005. Metric analysis of sex differences in South African black and white pelves. *Forensic Science International* 147: 119-127.

Penney AE. 2005. Analysis of the Talus and Calcaneus bones from the Poole-Rose Ossuary: A Late Woodland Burial Site in Ontario, Canada. *Unpublished MA thesis, Louisiana State University, Louisiana, U.S.A.*

Phenice TW. 1969. A newly developed visual method of sexing the os pubis. *American Journal of Physical Anthropology* 30: 297-301.

Plochocki JH. 2011. Sexual dimorphism of anterior sacral curvature. *Journal of Forensic Sciences* 56: 161-164.

Polednak AP. 1989. *Racial and ethnic differences in disease*. Oxford: Oxford University Press.

Pomeroy E, Zakrzewski SR. 2009. Sexual dimorphism in diaphyseal cross-sectional shape in the Medieval Muslim population of Écija, Spain, and Anglo-Saxon Great Chesterford, UK. *International Journal of Osteoarchaeology* 19: 50-65.

Pommerol F. 1869. *Recherches sur la synostose des os du crâne*. Thesis, l'École de Médecine, Paris.

Powers R. 1962. The disparity between known age and age as estimated by cranial suture closure. *Man* 84: 52-54.

Pretorius E, Steyn M, Scholtz Y. 2006. Investigation into the usability of geometric morphometric analysis in assessment of sexual dimorphism. *American Journal of Physical Anthropology* 129: 64-70.

Purkait R. 2005. Triangle identified at the proximal end of femur: a new sex determinant. *Forensic Science International* 147: 135-139.

Purkait R, Chandra H. 2004. A study of sexual variation in Indian femur. *Forensic Science International* 146: 25-33.

Quigley C. 2001. *Skulls and Skeletons: Human Bone Collections and Accumulations*. McFarland & Company, Inc., Publishers: North Carolina.

Relethford JH. 2002. Apportionment of global human genetic diversity based on craniometrics and skin color. *American Journal of Physical Anthropology* 118: 393-398.

Relethford JH. 2009. Race and global patterns of phenotypic variation. *American Journal of Physical Anthropology* 139: 16-22.

Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* 38: 735-742.

Ricklefs RE. 2008. The evolution of senescence from a comparative perspective. *Functional Ecology* 22: 379-392.

Rios L, Weisensee K, Rissech C. 2008. Sacral fusion as an aid in age estimation. *Forensic Science International* 180: 111.e1-111.e7.

Rissech C, Estabrook GF, Cunha E, Malgosa A. 2006. Using the acetabulum to estimate age-at-death of adult males. *Journal of Forensic Sciences* 51: 213-229.

Rissech C, Estabrook GF, Cunha E, Malgosa A. 2007. Estimation of age-at-death for adult males using the acetabulum, applied to four Western European populations. *Journal of Forensic Sciences* 52: 774-778.

Rissech C, García M, Malgosa A. 2003. Sex and age diagnosis by ischium morphometric analysis. *Forensic Science Inernational* 135: 188-196.

Rissech C, Malgosa A. 2005. Ilium growth study: applicability in sex and age diagnosis. *Forensic Science International* 147: 165-174.

Rissech C, Malgosa A. 2007. Pubis growth study: applicability in sexual and age diagnostic. *Forensic Science International* 173: 137-145.

Rissech C, Schaefer M, Malgosa A. 2008. Development of the femur – implications for age and sex determination. *Forensic Science International* 180: 1-9.

Rissech C, Steadman DW. 2011. The demographic, socio-economic and temporal contextualisation of the Universitat Autònoma de Barcelona Collection of Identified Human Skeletons (UAB Collection). *International Journal of Osteoarchaeology* 21: 313-322.

Roberts CA. 2007. A bioarchaeological study of maxillary sinusitis. *American Journal of Physical Anthropology* 133: 792-807.

Roberts DF. 1981. Genetics of growth. *British Medical Bulletin*. 37: 239-246.

Roberts CA, Manchester K. 2007. *The Archaeology of Disease*. Third Edition. Cornell University Press: New York.

Roger VL, Go AS, Lloyd-Jones DM, Adams RJ, Berry JD, Brown TM, Carnethon MR, Dai S, de Simone G, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Greenlund KJ, Hailpern SM, Heit JA, Ho PM, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, McDermott MM, Meigs JB, Moy CS,, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Rosamond WD, Sorlie PD, Stafford RS, Turan TN, Turner MB, Wong ND, Wylie-Rosett J. 2011. Heart disease and stroke statistics – 2011 update: a report from the American Heart Association. *Circulation* 123: e18-e209.

Rogers TL. 1999. A visual method of determining the sex of skeletal remains using the distal humerus. *Journal of Forensic Sciences* 44: 57-60.

Rogers TL. 2005. Determining the sex of human remains through cranial morphology. *Journal of Forensic Sciences* 50: 493-500.

Rogers TL. 2009. Sex determination of adolescent skeletons using the distal humerus. *American Journal of Physical Anthropology* 140: 143-148.

Rogers TL, Saunders S. 1994. Accuracy of sex determination using morphological traits of the human pelvis. *Journal of Forensic Sciences* 39: 1047-1056.

Rosas A, Pérez PJ, Bone J. 1999. Senescence in European Middle Pleistocene hominids: the Atapuerca evidence. *Human Evolution* 14: 83-98.

Rosenberg K. 2002. A late Pleistocene human skeleton from Liujiang, China suggests regional population variation in sexual dimorphism in the human pelvis. *Evolution and Variation* 10: 5-17.

Rougé-Maillart C, Vielle B, Jousset N, Chappard D, Telmon N, Cunha E. 2009. Development of a method to estimate skeletal age-at-death in adults using the acetabulum and the auricular surface on a Portuguese population. *Forensic Science International* 188: 91-95.

Rowe JW, Kahn RL. 1987. Human aging: usual and successful. *Science* 237: 143-149.

Ruff C. 1987.Sexual dimorphism in human lower limb bone structure: relationship to subsistence strategy and sexual division of labor. *Journal of Human Evolution* 16: 391-416.

Ruff CB, Hayes WC. 1982. Subperiosteal expansion and cortical remodeling of the human femur and tibia with aging. *Science* 217: 945-948.

Rühli FJ, Hotz G, Böni T. 2003. Brief communication: the Galler Collection. A little-known historic Swiss bone pathology reference series. *American Journal of Physical Anthropology* 121: 15-18.

Russell KF, Simpson SW, Genovese J, Kinkel MD, Meindl RS, Lovejoy CO. 1993. Independent test of the fourth rib aging technique. *American Journal of Physical Anthropology* 92: 53-62.

Sabini RC, Elkowitz DE. 2006. Significance of differences in patency among cranial sutures. *Journal of the American Osteopathic Association* 106: 600-604.

Sahni D, Jit I, Neelam S. 2005. Time of closure of cranial sutures in northwest Indian adults. *Forensic Science International* 148: 199-205.

Sahni S, Zoltick ES, McLean RR, Hannan MT. 2010. Non-D vitamins and bone health in adults. *IBMS BoneKEy* 7: 431-446.

Sakaue K. 2006. Application of the Suchey-Brooks system of pubic age estimation to recent Japanese skeletal material. *Anthropological Science* 114: 59-64.

Sakuma A, Ohtani S, Saitoh H, Iwase H. 2012. Comparative analysis of aspartic acid racemisation methods using whole-tooth and dentin samples. *Forensic Science International* 223: 198-201.

Salter DM, Lee H-S. 2012. Pathology and risk factors in osteoarthritis. *Journal of Medical Sciences* 32: 199-207.

Samworth R, Gowland R. 2007. Estimation of adult skeletal age-at-death: statistical assumptions and applications. *International Journal of Osteoarchaeology* 17: 174-188.

San Millán M, Rissech C, Turbón D. 2013. A test of Suchey-Brooks (pubic symphysis) and Buckberry-Chamberlain (auricular surface) methods on an identified Spanish sample: paleodemographic implications. *Journal of Archaeological Science* 40: 1743-1751.

Santori C, Ceccanti M, Diacinti D, Attilia ML, Toppo L, D'Erasmo E, Romagnoli E, Mascia ML, Cipriani C, Prastaro A, Carnevale V, Minisola S. 2008. Skeletal turnover, bone mineral density, and fractures in male chronic abusers of alcohol. *The Journal of Endocrinological Investigation* 31: 321-326.

Santos AL. 2000. A Skeletal Picture of Tuberculosis. Macroscopic, Radiological, Biomolecular, and Historical Evidence from the Coimbra Identified Skeletal Collection. *Unpublished PhD thesis, University of Coimbra.*

Santos AL, Roberts CA. 2001. A picture of tuberculosis in young Portuguese people in the early 20th century: a multidisciplinary study of the skeletal and historical evidence. *American Journal of Physical Anthropology* 115: 38-49.

Santos AL, Roberts CA. 2006. Anatomy of a serial killer: differential diagnosis of tuberculosis based on rib lesions of adult individuals from the Coimbra Identified Skeletal Collection, Portugal. *American Journal of Physical Anthropology* 130: 38-49.

Sashin D. 1930. A critical analysis of the anatomy and the pathologic changes of the sacro-iliac joints. *The Journal of Bone and Joint Surgery* 12: 891-910.

Saunders SR, Chan AHW, Kahlon B, Kluge HF, FitzGerald CM. 2007. Sexual dimorphism of the dental tissues in human permanent mandibular canines and third premolars. *American Journal of Physical Anthropology* 133: 735-740.

Saunders SR, DeVito C. 1991. Subadult skeletons in the Raymond Dart Anatomical Collection: research potential. *Human Evolution* 6: 421-434.

Saunders SR, Fitzgerald C, Rogers T, Dudar C, McKillop H. 1992. A test of several methods of skeletal age estimation using a documented archaeological sample. *Canadian Society of Forensic Science Journal* 25: 97-118.

Saunders SR, Herring A, Sawchuk LA, Boyce G. 1995. The nineteenth-century cemetery at St. Thomas' Anglican Church , Belleville: skeletal remains, parish records and censuses. In *Grave Reflections: Portraying the Past Through Cemetery Studies*, Saunders SR, Herring A (eds.). Canadian Scholars Press: Toronto; 93-117.

Sauvage HE. 1870. Note sur l'état sénile du crâne. *Bulletins de la Société d'Anthropologie de Paris*, Series II, Tome 5: 576-586.

Saville PD. 1965. Changes in bone mass with age and alcoholism. *Journal of Bone and Joint Surgery* 47: 492-499.

Sayer A, Cooper C, Evans JR, Rauf A, Wormald RPL, Osmond C, Barker DJP. 1998. Are rates of ageing determined *in utero*? *Age and Ageing* 27: 579-583.

Scheidel W. 1996. What's in an age? A comparative view of bias in the census returns of Roman Egypt. *Bulletin of the American Society of Papyrologists* 33: 25-59.

Scheidel W. 2001a. Progress and problems in Roman demography. In *Debating Roman Demography*, Scheidel W (ed.). Brill: Boston; 1-82.

Scheidel W. 2001b. Roman age structure: evidence and models. *The Journal of Roman Studies* 91: 1-26.

Scheuer L. 2002. Application of osteology to forensic medicine. *Clinical Anatomy* 15: 297-312.

Scheuer L, Black S. 2000. *Developmental Juvenile Osteology*. Elsevier, Ltd.: London.

Scheuer L, Bowman JE. 1995. Correlation of documentary and skeletal evidence in the St. Bride's crypt population. In *Grave Reflections: Portraying the Past Through Cemetery Studies*, Saunders S, Herring A (eds.). Canadian Scholars Press, Inc: Toronto; 49-70.

Scheuer L, MacLaughlin-Black S. 1994. Age estimation from the pars basilaris of the fetal and juvenile occipital bone. *International Journal of Osteoarchaeology* 4: 377-380.

Schmitt A. 2004. Age-at-death assessment using the os pubis and the auricular surface of the ilium: a test on an identified Asian sample. *International Journal of Osteoarchaeology* 14: 1-6.

Schmitt A, Murail P, Cunha E, Rougé D. 2002. Variability of the pattern of aging on the human skeleton: evidence from bone indicators and implications on age-at-death estimation. *Journal of Forensic Sciences* 47: 1-7.

Schmitt A, Saliba-Serre B, Tremblay M, Martrille L. 2010. An evaluation of statistical methods for the determination of age of death using dental root translucency and periodontosis. *Journal of Forensic Sciences* 55: 590-596.

Schulz-Aellen M-F. 1997. *Aging and Human Longevity*. Birkhäuser: Berlin.

Scott EC. 1979. Dental wear scoring technique. *American Journal of Physical Anthropology* 51: 213-218.

Semine AA, Damon A. 1975. Costochondral ossification and aging in five populations. *Human Biology* 47: 101-116.

Sharman J. 2004. Sex determination using the clavicle: the Grant Collection. Paper presented at the Canadian Association for Physical Anthropology Meetings, London, Ontario.

Sheridan SG, Van Gerven DP. 1997. Female biological resiliency: differential stress response by sex in human remains from ancient Nubia. *Human Evolution* 12: 241-252.

Sherman S. 1999. Human aging at the millennium. In *The Aging Skeleton*, Rosen CJ, Glowacki J, Bilezikian JP (eds.). Academic Press: London; 11-18.

Silverberg SJ, Bilezikian JP. 1999. Parathyroid hormone. In *The Aging Skeleton*, Rosen CJ, Glowacki J, Bilezikian JP (eds.). Academic Press: London; 175-184.

Singh S, Potturi BR. 1978. Greater sciatic notch in sex determination. *Journal of Anatomy* 125: 619-624.

Smith BH, Garn SM, Hunter WS. 1986. Secular trends in face size. *The Angle Orthodontist* 56: 196-204.

Spirduso WW. 1995. *Physical Dimensions of Aging*. Human Kinetics: Champaign, IL.

St. Hoyme LE, İşcan MY. 1989. Determination of sex and race: accuracy and assumptions. In *Reconstruction of Life from the Skeleton*, İşcan MY, Kennedy KAR (eds.). Alan R. Liss, Inc.: New York; 53-93.

Stanton SJ, Liening SH, Schultheiss OC. 2011. Testosterone is positively associated with risk taking in the Iowa Gambling Task. *Hormones and Behavior* 59: 252-256.

Star H, Thevissen P, Jacobs R, Fieuws S, Solheim T, Willems G. 2011. Human dental age estimation by calculation of pulp-tooth volume ratios yielded on clinically acquired cone beam computer tomography images of monoradicular teeth. *Journal of Forensic Sciences* 56: S77-S82.

Steadman DW, Adams BJ, Konigsberg LW. 2006. Statistical basis for positive identification in forensic anthropology. *American Journal of Physical Anthropology* 131: 15-26.

Steele DG. 1976. The estimation of sex on the basis of the talus and calcaneus. *American Journal of Physical Anthropology* 45: 581-588.

Stevenson JC, Mahoney ER, Walker PL, Everson PM. 2009. Technical note: prediction of sex based on five skull traits using decision analysis (CHAID). *American Journal of Physical Anthropology* 139: 434-441.

Stewart TD. 1954. Sex determination of the skeleton by guess and by measurement. *American Journal of Physical Anthropology* 12: 385-392.

Stewart TD. 1984. Pathologic changes in aging sacroiliac joints: a study of dissecting-room skeletons. *Clinical Orthopaedics and Related Research* 183: 188-196.

Steyn M, İşcan MY. 1997. Sex determination from the femur and tibia in South African whites. *Forensic Science International* 90: 111-119.

Steyn M, İşcan MY. 1998. Sexual dimorphism in the crania and mandibles of South African whites. *Forensic Science International* 98: 9-16.

Steyn M, İşcan MY. 1999. Osteometric variation in the humerus: sexual dimorphism in South Africans. *Forensic Science International* 106: 77-85.

Steyn M, Pretorius E, Hutten L, 2004. Geometric morphometric analysis of the greater sciatic notch in South Africans. *HOMO – Journal of Comparative Human Biology* 54: 197-206.

Stini WA. 1969. Nutritional stress and growth: sex difference in adaptive response. *American Journal of Physical Anthropology* 31: 417-426.

Stini WA. 1994. Comparisons of Chinese and Arizona bone densities. *Collegium Antropologicum* 18: 155-163.

Stinson S. 1985. Sex differences in environmental sensitivity during growth and development. *Yearbook of Physical Anthropology* 28: 123-147.

Storey R. 2007. An elusive paleodemography? A comparison of two methods for estimating the adult age distribution of deaths at Late Classic Copan, Honduras. *American Journal of Physical Anthropology* 132: 40-47.

Stott GG, Sis RF, Levy BM. 1982. Cemental annulations as an age criterion in forensic dentistry. *Journal of Dental Research* 61: 814-817.

Straus Jr. WL. 1927. The human ilium: sex and stock. *American Journal of Physical Anthropology* 11: 1-28.

Suchey JM. 1979. Problems in the aging of females using the *os pubis*. *American Journal of Physical Anthropology* 51: 467-470.

Suekoa KT, Ewing SK, Ensrud KE, Laughlin GA, Hoffman AR, Varosy PD. 2010. Higher endogenous testosterone levels associated with increased risk of coronary heart disease in elderly men. *Endocrine Reviews Supplement* 1: S858.

Sulzmann CE, Buckberry JL, Pastor RF. 2008. The utility of carpals for sex assessment: a preliminary study. *American Journal of Physical Anthropology* 135: 252-262.

Syme SL, Berkman LF. 1976. Social class, susceptibility and sickness. *American Journal of Epidemiology* 104: 1-8.

Tal H, Tau S. 1983. Statistical survey of the human skulls in the Raymond Dart Collection of skeletons. *South African Journal of Science* 79: 215-217.

Tal H, Tau S. 1984. Tooth loss and tooth retention in a multitribal group of Bantu-speaking South African blacks: a study on 500 dry mandibles. *American Journal of Physical Anthropology* 64: 75-82.

Thieme FP, Schull WJ. 1957. Sex determination from the skeleton. *Human Biology* 29: 242-273.

Thomas, DH. 1986. Refiguring anthropology – first principles of probability and statistics. Waveland Press, Inc.: Prospect Heights.

Tobias P. 1988. War and the negative secular trend of South African blacks with observations on the relative sensitivity of cadaveric and non-cadaveric populations to secular effects. *Proceedings of the 5th Congress of the European Anthropological Association* I, Lisbon; 451-461.

Tobias PV. 1987. Memories of Robert James Terry (1871-1966) and the genesis of the Terry and Dart collections of human skeletons. *Adler Museum Bulletin* 13: 31-34.

Tobias PV. 1991. On the scientific, medical, dental and educational value of collections of human skeletons. *International Journal of Anthropology* 6: 277-280.

Todd TW. 1920. Age changes in the pubic bone. I. The male white pubis. *American Journal of Physical Anthropology* 3: 285-339.

Todd TW. 1921. Age changes in the pubic bone. II. The pubis of the male negro-white hybrid. *American Journal of Physical Anthropology* 4: 1-25.

Todd TW. 1921b. Age changes in the pubic bone. III. The pubis of the white female. *American Journal of Physical Anthropology* 4: 26-39.

Todd TW. 1921c. Age changes in the pubic bone. IV. The pubis of the female negro-white hybrid. *American Journal of Physical Anthropology* 4: 40-70.

Todd TW. 1921d. Age changes in the pubic bone. V. Mammalian pubic metamorphosis. *American Journal of Physical Anthropology* 4: 333-406.

Todd TW. 1921e. Age changes in the pubic bone. VI. The interpretation of variations in the symphysial area. *American Journal of Physical Anthropology* 4: 407-424.

Todd TW. 1923. Age changes in the pubic symphysis. VII. The anthropoid strain in human pubic symphyses of the third decade. *Journal of Anatomy* 57: 274-294.

Todd TW. 1930. Age changes in the pubic bone. VIII. Roentogenographic differentiation. *American Journal of Physical Anthropology* 14: 255-271.

Todd TW, Lyon DW. 1924. Endocranial suture closure. Its progress and age relationship. Part I. Adult males of white stock. *American Journal of Physical Anthropology* 7: 325-384.

Todd TW, Lyon DW. 1925. Cranial suture closure. Its progress and age relationship. Part II. Ectocranial closure in adult males of white stock. *American Journal of Physical Anthropology* 8: 23-45.

Todd TW, Lyon DW. 1925b. Cranial suture closure. Its progress and age relationship. Part III. Endocranial closure in adult males of negro stock. *American Journal of Physical Anthropology* 8: 47-71.

Todd TW, Lyon DW. 1925c. Cranial suture closure. Its progress and age relationship. Part IV. Ectocranial closure in adult males of negro stock. *American Journal of Physical Anthropology* 8: 149-168.

Toivanen AT, Heliōvaara M, Impivaara O, Arokoski JPA, Knekt P, Lauren H, Krōger H. 2010. Obesity, physically demanding work and traumatic knee injury are major risk factors for knee osteoarthritis – a population-based study with a follow-up of 22 years. *Rheumatology* 49: 308-314.

Tortella G. 1994. Patterns of economic retardation and recovery in south-western Europe in the nineteenth and twentieth centuries. *Economic History Review* 47: 1-21.

Towfighi A, Saver JL. 2011. Stroke declines from third to fourth leading cause of death in the United States: historical perspectives and challenges ahead. *Stroke* 42: 2351-2355.

Trotter M, Broman GE, Peterson RR. 1959. Density of cervical vertebrae and comparison with densities of other bones. *American Journal of Physical Anthropology* 17: 19-25.

Trudell MB. 1999. Anterior femoral curvature revisited: race assessment from the femur. *Journal of Forensic Sciences* 44: 700-707.

Turner RT. 2000. Skeletal response to alcohol. *Alcoholism:Clinical and Experimental Research* 24: 1693-1701.

Turner W. 1885. The index of the pelvic brim as a basis of classification. *Journal of Anatomy and Physiology* 20: 125-143.

Usher BM. 2002. Reference samples: the first step in linking biology and age in the human skeleton. In *Paleodemography – Age Distributions from Skeletal Samples*, Hoppa RD, Vaupel JW (eds.). Cambridge University Press: Cambridge; 29-47.

Usher BM. Website: < http://www2.potsdam.edu/usherbm/reference/>

Uytterschaut HT. 1986. Sexual dimorphism in human skulls. A comparison of sexual dimorphism in different populations. *Human Evolution* 1: 243-250.

Van der Merwe AE, İşcan MY, L'Abbé EN. 2006. The pattern of vertebral osteophyte development in a South African population. *International Journal of Osteoarchaeology* 16: 459-464.

Van Gerven DP. 1972. The contribution of size and shape variation to patterns of sexual dimorphism of the human femur. *American Journal of Physical Anthropology* 37: 49-60.

Van Gerven DP, Armelagos GJ. 1983. "Farewell to paleodemography?" Rumors of its death have been greatly exaggerated. *Journal of Human Evolution* 12: 353-360.

Veugelers PJ, Yip AM, Kephart G. 2001. Proximate and contextual socioeconomic determinants of mortality: multilevel approaches in a setting with universal health care coverage. *American Journal of Epidemiology* 154: 725-732.

Vlak D, Roksandic M, Schillaci MA. 2008. Greater sciatic notch as a sex indicator in juveniles. *American Journal of Physical Anthropology* 137: 309-315.

Wade A, Nelson A, Garvin G, Holdsworth DW. 2011. Preliminary radiological assessment of age-related change in the trabecular structure of the human *os pubis*. *Journal of Forensic Sciences* 56: 312-319.

Waldron HA. 1997. Association between osteoarthritis of the hand and knee in a population of skeletons from London. *Annals of the Rheumatic Diseases* 56: 116-118.

Waldron T. 2009. *Palaeopathology*. Cambridge Manuals in Archaeology. Cambridge University Press: Cambridge.

Waldron HA, Cox M. 1989. Occupational arthropathy: evidence from the past. *British Journal of Industrial Medicine* 46: 420-422.

Walker PL, Johnson JR, Lambert PM. 1988. Age and sex biases in the preservation of human skeletal remains. *American Journal of Physical Anthropology* 76: 183-188.

Walker PL. 1995. Problems of preservation and sexism in sexing: some lessons from historical collections and palaeodemographers. In *Grave Reflections: Portraying the Past Through Cemetery Studies*, Saunders SR, Herring A (eds.). Canadian Scholars Press: Toronto; 31-47.

Walker PL. 2005. Greater sciatic notch morphology: sex, age, and population differences. *American Journal of Physical Anthropology* 127: 385-391.

Walker PL. 2008. Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology* 136: 39-50.

Walker RA, Lovejoy CO. 1985. Radiographic changes in the clavicle and proximal femur and their use in the determination of skeletal age-at-death. *American Journal of Physical Anthropology* 68: 67-78.

Walrath DE, Turner P, Bruzek J. 2004. Reliability test of the visual assessment of cranial traits for sex determination. *American Journal of Physical Anthropology* 125: 132-137.

Washburn SL. 1948. Sex differences in the pubic bone. *American Journal of Physical Anthropology* 6: 199-207.

Washburn SL. 1949.Sex differences in the pubic bone of Bantu and Bushman. *American Journal of Physical Anthropology* 7: 425-432.

Wasterlain SN, Hillson S, Cunha E. 2009. Dental caries in a Portuguese identified skeletal sample from the late 19th and early 20th centuries. *American Journal of Physical Anthropology* 140: 64-79.

Watanabe S, Terazawa K. 2006. Age estimation from the degree of osteophyte formation of vertebral columns in Japanese. *Legal Medicine* 8: 156-160.

Weisensee KE, Jantz RL. 2011. Secular changes in craniofacial morphology of the Portuguese using geometric morphometrics. *American Journal of Physical Anthropology* 145: 548-559.

Weise S, Boldsen JL, Gampe J, Milner GR. 2009. Calibrated Expert Inference and the construction of unbiased paleodemographic mortality profiles. (Abstract) Paper presented at the 78th Annual Meeting of the American Association of Physical Anthropologists. *American Journal of Physical Anthropology* 138: 269.

Weiss KM. 1972. On the systematic bias in skeletal sexing. *American Journal of Physical Anthropology* 37: 239-250.

Weiss E, Jurmain R. 2007. Osteoarthritis revisited: a contemporary review of aetiology. *International Journal of Osteoarchaeology* 17: 437-450.

Weon BM, Je JH. 2009. Theoretical estimation of maximum human lifespan. *Biogerontology* 10: 65-71.

Williams BA, Rogers TL. 2006. Evaluating the accuracy and precision of cranial morphological traits for sex determination. *Journal of Forensic Sciences* 51: 729-735.

Wilson LA, MacLeod N, Humphrey LT. 2008. Morphometric criteria for sexing juvenile human skeletons using the ilium. *Journal of Forensic Sciences* 53: 269-278.

Wilson RJ, Algee-Hewitt B, Jantz LM. 2007. Demographic trends within the Forensic Anthropology Center's body donation program. Annual Meeting of American Association of Physical Anthropologists, abstract. *American Journal of Physical Anthropology Supplement* (44) 132: 252.

Wiredu E, Kumoji R, Seshadri R, Biritwum RB. 1999. Osteometric analysis of sexual dimorphism in the sternal end of the rib in a West African population. *Journal of Forensic Sciences* 44: 921-925.

Wittwer-Backofen U, Buckberry J, Czarnetzki A, Doppler S, Grupe G, Hotz G, Kemkes A, Larsen CS, Prince D, Wahl J, Fabig A, Weise S. 2008. Basics in paleodemography: a comparison of age indicators applied to the early medieval skeletal sample of Lauchheim. *American Journal of Physical Anthropology* 137: 384-396.

Wolff K, Vas Z, Sótonyi P, Magyar LG. 2012. Skeletal age estimation in Hungarian population of known age and sex. *Forensic Science International* 223: 374.e1-374.e8.

Wood JW, Milner GR, Harpending HC, Weiss KM. 1992. The osteological paradox. Problems of inferring prehistoric health from skeletal samples. *Current Anthropology* 33: 343-370.

World Health Organization. 2006. Country Health System Fact Sheet, 2006: South Africa. Online, http://www.afro.who.int/home/countries/fact_sheets/southafrica.pdf (accessed on 30/12/2009).

World Health Organization, 2011. The 10 leading causes of death by broad income group (2008). Fact sheet number 310. Updated June 2011. Online, http://www.who.int/mediacentre/factsheets/fs310/en/ (accessed on 05/03/2013).

Wright ND, Bahrami B, Johnson E, Di Malta G, Rees G, Frith CD, Dolan RJ. 2012. Testosterone disrupts human collaboration by increasing egocentric choices. *Proceedings of the Royal Society Biological Sciences* 279: 2275-2280.

Yavuz MF, İşcan MY, Çöloğlu AS. 1998. Age assessment by rib phase analysis in Turks. *Forensic Science International* 98: 47-54.

Yoder C, Ubelaker DH, Powell JF. 2001. Examination of variation in sternal rib end morphology relevant to age assessment. *Journal of Forensic Sciences* 46: 223-227.

Zak I, Jagielski T, Kwiatkowski S, Bielecki J. 2012. *Prototheca wickerhamii* as a cause of neuroinfection in a child with congenital hydrocephalus. First case of human protothecosis in Poland. *Diagnostic Microbiology and Infectious Disease* 74: 186-189.

Zhao Z. 1995. Record longevity in Chinese history – evidence from the Wang Genealogy. In *Exceptional Longevity: From Prehistory to the Present*, Odense Monographs on Population Aging 2, Jeune B, Vaupel JW (eds.). Odense University Press: Odense; 93-107.